

Drug safety on trial

The current US system for checking the safety of drugs already on the market is toothless. Why isn't the government doing more to strengthen it?

A revealing notice appeared last month in the *Federal Register*, the US government compendium of agency rules and notices. The Food and Drug Administration (FDA) was reporting on compliance by pharmaceutical companies with its requests for studies of the clinical safety and efficacy of drugs already on the market. Of nearly 1,200 such studies committed to by drug firms but not completed, some 70% have yet to begin.

This is an alarming reflection of the state of the government's vigilance. Last year, the FDA belatedly faced up to research showing that the painkiller Vioxx, which it approved in 1999, markedly increases the risk of heart attacks and strokes. It has been estimated that more than 25,000 people died before Merck pulled the drug from the market in September (see pages 554 and 557). A better approach to assessing the safety of marketed drugs is badly needed.

Under the FDA's existing system, known as MedWatch, doctors voluntarily report suspected side-effects — but epidemiologists estimate that this captures only 10% of adverse events. The agency used to complement this with university-run studies, but this modest effort dried up in the 1990s when it was forced to devote more resources to speeding up drug approvals.

The upshot is that the FDA depends on companies for post-market safety studies but has no legal authority to force firms to do them. The results are sobering. For instance, after 16 years of use in tens of millions of people, it is still not known whether selective serotonin

reuptake inhibitors, the blockbuster antidepressants, cause an increase in suicide attempts in some adults. The FDA warned last year of such a risk in young people, but the question still hasn't been adequately explored in adults. Nor will it be if the current system prevails.

Lack of information is not the problem. After a drug comes into use, gigabytes of epidemiological data become available from governments and healthcare organizations. Nor is it a matter of sorting through the 10,000-odd prescription drugs to find one with dangerous side-effects: the most likely culprits are already well known.

What is lacking is both the money and the mandate. The United States needs a government body that can not only determine what drug safety studies are needed, but demand that they are done. It must also have the authority to dictate — rather than negotiate with drug-makers, as the FDA currently does — that beefed-up warning labels are used when evidence of new risks emerges. About \$300,000 annually would pay for a high-quality pharmaco-epidemiological study. Surely the richest country in the world can find the funds for that.

It can, but will it? Congress is unlikely to implement such a change. The Republicans control Congress and the White House, and are loath to alienate an industry that has given them twice as much as it has given the Democrats in political contributions over the past decade.

Tens of thousands of people have almost certainly died because of Vioxx. Observers are left to wonder if it will take an even bigger tragedy to force the US government to do the right thing on drug safety. ■

A state of ignorance

Severe brain damage attracts little research attention, yet science could help inform the decisions of doctors and families.

The bitter wrangle over the fate of Terri Schiavo, the severely brain-damaged woman at the centre of a political and emotional storm in the United States, is highly distressing. Her husband has fought to allow her to die; her parents have argued that she might yet recover. As *Nature* went to press, Schiavo's parents seemed to have exhausted their legal options, and their daughter has had her feeding tube removed.

It is not our intention to pass judgment on Schiavo's fate. But for patients with related conditions, and their families and doctors, science may have a role to play that is not being fully explored.

In the case of Schiavo, whose brain was temporarily starved of oxygen 15 years ago, medical opinion seems clear. Neurologists say that she is in a persistent vegetative state and is unable to respond to instructions. Her chances of recovery are close to zero.

But when the prognosis is less certain, research may be able to help. Some severely brain-damaged patients are described as being in a minimally conscious state. These patients are occasionally able to respond to commands and are thought to have a slightly better chance of some recovery.

In a study published last month (N. D. Schiff *et al. Neurology* 64, 514–523; 2005), researchers probed the brains of two such patients using functional magnetic resonance imaging while relatives read them personal stories. The researchers found activity in language

networks that was similar to that of healthy individuals. Other scans have shown the brain regions that are physically damaged, but not which circuits can respond to stimuli.

The results are only preliminary: it is not clear, for example, what the patients actually experienced. But researchers could potentially distinguish a persistent vegetative state from a minimally conscious state, or even identify conditions in between. They might even find signature patterns that can help predict the likelihood of recovery.

Perhaps the biggest obstacle is the perception in the medical and research community that severely brain-damaged patients are a lost cause. This is reflected in the number of research groups scanning the brains of such patients; they can be counted on one hand.

Some researchers have to battle for permission to work with brain-damaged patients. Patients who cannot consent to research must clearly be safeguarded. But the situation could be simplified through living wills, which allow people to specify the treatment they want if they become incapacitated. These could include requests about participation in research, as some already do about organ donation.

The ethical and medical issues involved in these cases will always be excruciatingly difficult, even when the diagnoses are as clear-cut as Schiavo's. An increased motivation to tackle research in this area won't make the problems go away, but it could help to inform the difficult decisions that families and doctors are forced to take. ■



Corn dogged
Antibiotic genes
add to transgenic
crop scandal
p548



Page proofs
Journals lack
policies about
ethics of adverts
p549



Money worries
Vienna lab chief
charged with not
spending cash
p550



Sky hooks
NASA spurs
inventors to new
heights with prizes
p552

Millennium group nails down the financial value of ecosystems

Jim Giles, London

Ecosystems will continue to decline unless policy-makers start to assess the economic benefits of our natural environment. That's the verdict of a group that has just completed the most comprehensive round-up yet of the planet's ecological health.

The four-year Millennium Ecosystem Assessment was due to issue its summary report on 30 March in London and Washington. It makes for gloomy reading, stating that damage to ecosystems is irreversible, likely to accelerate over the next 50 years, and set to frustrate efforts to eradicate poverty.

The US\$24-million project brought together 1,300 biological, physical and social scientists from 95 countries. The researchers conclude that ecological threats can only be held in check if governments start to assign proper economic value to the benefits they obtain from natural systems (see commentary on pages 561–562).

Such 'ecosystem services' include products that governments already quantify — such as fishery income — and others that are taken for granted — the protection forests give against flooding, for example. The authors argue that less tangible benefits, such as aesthetic values, should also be considered.

"The idea is to take the hand-waving out of ecosystem decisions," says Robert Scholes, a specialist in environmental policy at the Council for Scientific and Industrial Research



World service: income from fisheries is an easily quantified benefit of a healthy ecosystem.

in Pretoria, South Africa, and a member of the project's scientific panel.

The assessment team used its ecosystem services concept to examine how Earth's environment will develop under four scenarios. In each case, ecosystem damage makes it difficult to meet a projected 70–80% increase in food demand over the next 50 years. Habitat loss always hits biodiversity, with the total number of plant species reduced by 10–15% of 1970 levels by 2050.

"It's not hopeless," says Scholes. "There is a large difference between bad and really bad." The authors suggest taking measures such as improving regulation and market incentives. Although these are not new ideas, the researchers claim they have enough

backing to make a difference. "We involved members of the United Nations, the scientific community, the private sector and indigenous peoples' groups," says Thomas Rosswall, head of the International Council for Science in Paris and a member of the board that managed the project.

But critics say the authors fail to note the benefits that richer nations have derived from activities that degrade ecosystems, such as felling forests. "This is how much of the West developed," says Bjørn Lomborg, a statistician at the University of Aarhus in Denmark and a high-profile critic of the Green movement. "We grew rich in the process and only then have we been able to reforest." Lomborg says that poorer nations should exploit ecosystems, and expect to address environmental concerns only once they have dealt with problems such as malnutrition.

The data produced by the project are now being distilled into special reports for five UN bodies, including the Convention on Biological Diversity. The World Bank intends to ask countries to use the concept of ecosystem services when submitting reports.

The assessment might even persuade countries that economic indicators are not the best way of describing a nation's well-being, says Robert May, president of the Royal Society in London and another panel member. "We need more awareness that GDP can't quantify quality of life and sustainability." ■

Indonesia spared tsunami as disaster quake strikes

Michael Hopkin, London

A huge, magnitude-8.7 earthquake shook Indonesia on 28 March, leaving 2,000 people feared dead on the island of Nias. But the event did not spark a tsunami like that of the 26 December earthquake, which occurred a little farther north.

Seismologists had recently concluded that the unruptured faults close to the 26 December epicentre, including the one that caused this earthquake, were close to failure (see pages 573–574). The latest event could

verify such work. "This quake may be a vindication of the 'stress transfer process,'" says Bill McGuire, an expert on earthquake hazards at University College London.

As *Nature* went to press, seismologists were still working out why the huge undersea earthquake, which apparently lasted some two minutes, did not set off a wave. The earthquake, among the eight largest on the planet since 1900, occurred in an area where lesser events have triggered killer waves before. One possibility is that

the earthquake occurred deeper in the Earth than its December predecessor, and so did not shunt up the sea floor abruptly.

Thousands of people in Sumatra, Thailand and Malaysia fled the coasts as they felt the shaking on Monday, and governments issued a tsunami alert using radio and television broadcasts. The area is due to get a dedicated tsunami warning system by the end of this year, which should improve evacuation procedures and prevent false alarms. ■

Stray seeds had antibiotic-resistance genes

Colin Macilwain, Washington

Hundreds of tonnes of genetically modified corn seeds sold to farmers by mistake over the past four years contained a gene for antibiotic resistance, *Nature* has learned. The release of such genes into the environment is sometimes considered inadvisable, as there is a small chance that they could flow from crops to microorganisms and spread problems of antibiotic resistance.

The Swiss biotechnology company Syngenta admitted last week that it had accidentally released a variety of corn (maize) called *Bt10* between 2001 and 2004. Like other crops with the name *Bt*, this corn had been genetically modified to produce a protective pesticide. But *Bt10* has not been approved for sale by regulatory agencies.

Officials at the company last week argued that *Bt10* is basically identical to *Bt11* corn, which has been approved for sale (see *Nature* 434, 423; 2005). But this week, Sarah Hull, a spokeswoman for Syngenta, confirmed that a marker gene that confers resistance to ampicillin, a commonly used antibiotic, was present in the *Bt10* seeds. She adds that this gene would not have been active in the corn plants that grew from the seeds.

Antibiotic-resistance genes are widely used as 'tags' during the production of genetically modified crops, to help breeders identify and preserve desirable strains. But the genes are often removed before the seeds enter the food chain. The presence of the marker gene in *Bt10* corn was noted in a 2003 advice notice from a UK government committee, the Advisory Committee on Releases to the Environment, which was using *Bt10* as a comparison to prove that

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Keeping track: antibiotic-resistance genes are often used as markers during the production of transgenic crops such as *Bt* corn.

there were no marker genes in *Bt11* corn.

Critics have expressed surprise that neither Syngenta nor the US Environmental Protection Agency (EPA) announced the presence of the marker when they admitted that the release of *Bt10* had taken place. "It is quite scandalous," says Greg Jaffe, head of the biotechnology project at the Center for Science in the Public Interest, a pressure group in Washington DC. "This shows that the government and the company are not being forthright."

Hull says that the company didn't mention the gene's presence because "it wasn't relevant to the health and safety discussion". She adds that the antibiotic-resistance genes have been around for a long time. "They've been studied extensively, and they pose no risk to humans or animals," she says. Regulators say that the genes present a very small risk to

human health, either directly — if in the stomach of a patient on antibiotics, for example — or indirectly through gene flow into microbes.

Michael Rodemeyer, director of the Pew Initiative on Food and Biotechnology, a think-tank in Washington DC, says that the presence of such genes would be unlikely to see a crop declared unsafe in the United States — but adds that it could cause problems in Europe.

In a ruling published last April, for example, the European Food Safety Authority, which advises European Union governments on food issues, said that marker genes conferring resistance to ampicillin "should be restricted to field trials and not be present in genetically modified plants placed on the market". And the Codex Alimentarius Commission, the international food-standards body, has urged the agricultural biotechnology industry to use alternative methods to refine genetically modified strains in the future.

The EPA, which is jointly investigating the release of the *Bt10* corn with the US Department of Agriculture, declined to say what it knew about the antibiotic-resistance marker. "What the company told us and when about the marker gene is part of our ongoing investigation and we are not able to discuss it at this time," the agency said in a statement.

"I think they've done a terrible job," says Margaret Mellon, head of the food and environment programme at the Union of Concerned Scientists in Washington DC, referring to both Syngenta and the government agencies. "There are lots and lots of unanswered questions, and the longer they remain, the less confidence people are going to have in the technology and in the regulatory system." ■

Deal paves way for Congress vote on stem-cell rules

Erika Check, Washington

The US Congress looks set to vote on whether to loosen the restrictions that currently govern human embryonic stem-cell research.

Under an agreement reached earlier this month, the Republican leadership of the House of Representatives will for the first time allow the body to debate and vote on whether to change the regulations.

The current rules, which were announced by President George W. Bush on 9 August 2001, prohibit the use of federal money to pay for research on human embryonic

stem-cell lines created after that date.

Congressman Mike Castle (Republican, Delaware), who favours easing the restrictions, says that he has reached a deal with Republican leaders Roy Blunt (Missouri), Eric Cantor (Virginia) and House speaker Dennis Hastert (Illinois). "My sense is that the speaker means to have a legitimate debate and a vote on whether to open up the policies on embryonic stem-cell research," Castle says.

Until now, the Republican leadership has opposed a review of Bush's rules. But developments since the policy was

announced have convinced some prominent Republicans, including Senator Orrin Hatch (Utah), to reconsider. Scientists have said that there are not as many cell lines available as the president initially said there would be. And the lines have all been made or grown on animal cells, which means that they will not be suitable for clinical trials in people.

Castle claims that a 100 million people could benefit from the research. "I think the leadership understands the importance of that," Castle says, "and frankly some of them are probably in favour of it quietly, if not openly." ■

Journals lack explicit policies for separating eds from ads

Jim Giles, London

Two in five biomedical journals have no declared policy on how to separate editorial and commercial matters, according to the preliminary results of a survey of editors.

Some journal editors say they find the results disturbing, given the commercial pressure that journals have to deal with. Many say they have experienced cases where companies have sought, to varying degrees, to influence journal content.

The survey asked the 350 members of the Committee on Publication Ethics (COPE), a London-based association of journal editors and publishers, whether their employers had a “declared policy on how to ensure a separation between editorial and commercial decisions”. The 118 responses received by 11 March indicate that 39% of respondents’ journals had no such policy.

Reprints are one area in which editorial and financial issues can become entangled, says Alex Williamson, publishing director at the *BMJ* (formerly *British Medical Journal*) Publishing Group in London. She has seen submission letters in which authors claimed that, should the paper involved be accepted, a pharmaceutical company would buy large numbers of reprints. “That can be very lucrative for the journal,” Williamson says.

Some editors argue that formal policies are required to help them fend off more subtle attempts from companies to influence their content. “It doesn’t have to be direct,” agrees Merrill Goozner, director of the Integrity in Science project at the Center for Science in the Public Interest in Washington DC. “It can be implied.”

And editors warn of the dangers of approaching firms with offers to sell adverts next to papers about the product being advertised, a practice that many publications have policies against. “I know other journals do this,” Williamson says. “That can put editors under pressure.”

Some journals, including the medical giants such as the *BMJ* and *The Lancet*, check the factual content of the adverts that they run. But smaller journals often lack the resources to make such checks, say COPE members. A senior staff member at one journal that

does check — who asked not to be named — says that he asks for changes to be made to one in every three adverts submitted to his publication. Monitoring often uncovers the use of references that do not back up the claims made in the advert, he says, adding that the process is “very labour intensive”.

Public rows between sponsors and journals may also put pressure on editors. In 2000, the house journal of the Hastings Center, a bioethics research institute in Garrison, New York, published a critical article on antidepressant drugs by David Healy, a director of psychological medicine at the University of Cardiff, Wales. A few months after publication, pharmaceutical manufacturer Eli Lilly withdrew its annual \$25,000 grant to the centre, saying that the Healy article was “inaccurate and unbalanced”.

Some editors fear that such incidents could make journals think twice before publishing material attacking pharmaceutical or other industries. To provide editors with support if they feel under pressure, COPE members suggest that journals produce policies that cover relationships with advertisers, sponsors and firms involved in any other revenue streams. The policies would stress that issues such as reprint sales should not influence the peer-review process and that adverts should be checked for accuracy.

But some publishing insiders suggest that the 39% figure may overstate the problem. “The policies may not be explicit, but they are almost a given at most organizations,” says Andrea Powell, chair of the Association of Learned and Professional Society Publishers, based in Worthing, West Sussex. “Editors are not likely to be swayed, as they don’t want their objectivity thrown into doubt.”

■ www.publicationethics.org.uk

WWF warns that China’s forests are not out of the woods

Rachael Williams, Tokyo

Government incentives to restore tree cover in China have damaged the environment and levied a high cost on forests worldwide, according to a WWF report issued this March. The critique suggests that China is shifting the ecological burden of its industrialization on to other countries.

China has been keen to impose logging bans and promote replanting schemes since floods in 1998 killed thousands of people and inundated millions of hectares of farmland. The conservation group WWF has largely approved of these policies, particularly as they expand the habitat of the endangered giant panda.

But, although official statistics show that the total amount of forest in China has increased by 1.2% a year over the past decade, much of the restored forest consists of single-species plantations on previously non-forested land, says the WWF report. As logging of some natural forest continues, this means that China’s forest diversity is probably declining, it says.

As its housing and energy needs grow, China continues to be one of the world’s largest importers of illegally logged wood. To fill the gap between supply and demand it is estimated that China will need to import 125 million cubic metres of wood in the year 2010. More than half of this will probably come from Russia, Indonesia and Malaysia, where logging is poorly regulated.

Illegal logging in such countries has long been seen as an important international issue. At an 18 March prelude to the G8 summit due to be held in Scotland this July, environmental and development ministers from the world’s eight leading industrialized nations committed to voluntary measures to address the problem. But environmental groups criticized the pledge, saying that legislative actions are needed. They added that countries outside the G8 — particularly China — also need to make responsible decisions about the resources they need.

David Kaimowitz, director of the Center for International Forestry Research, based in Indonesia, says it is China’s ultimate aim to get most of its forestry products from its own plantations. “But no one knows when this will happen, if ever,” he says. “For the sake of the world’s forests, let us hope it does.”



Book worms: advertisers can put editors under pressure.

Biologists snub 'kangaroo court' for Darwin

Geoff Brumfiel, Washington

Kansas biologists are set to boycott upcoming board of education hearings on the future of science teaching in the state.

The researchers contend that the hearings are being set up to serve as a thinly veiled showcase for 'intelligent design' — the theory that a god shaped the course of evolution.

Over six days in early May, the board hopes to hear arguments from proponents of intelligent design and scientists about whether there is evidence for the intervention of a deity in the process of evolution. "We view this hearing as an opportunity to educate the committee and the public," says Steve Abrams, a veterinarian in Arkansas City and chairman of the board of education. The hearings are part of a review of the state's science curriculum.

But so far, no evolutionary biologists have agreed to participate in the hearings. They say that the board has already decided to include language that is friendly to intelligent design in the new science standards. "We will not participate in their kangaroo

court," says Harry McDonald, president of Kansas Citizens for Science. "We will lose and the creationists will win if we lend our credibility to these hearings," he adds.

Kansas Citizens for Science is a group of pro-evolution researchers and science teachers that successfully opposed a 1999 drive to drop the teaching of evolution in the state (see *Nature* **400**, 701; 1999). The attempt outraged researchers and embarrassed state officials; in 2001 a newly elected school board voted to reaffirm the requirement to teach Darwin's theory.

But last November, conservatives regained a majority on the board, and are again considering revisions to the way evolution is taught. This time, however, the changes are more subtle. The revised plans were drawn up by a minority group on a 25-member panel that was appointed last June to write a science curriculum for Kansas.

The plan will introduce a definition of science that includes the possibility of the supernatural, and will point out several "weaknesses" of macroevolutionary theory,

such as gaps in the fossil record, says John Calvert, managing director of the Intelligent Design Network based in Shawnee, Kansas. "The idea is simply to open up the discussion of evolution," he says.

Abrams says the purpose of the hearings is to help educate board members about the proposed changes. But McDonald maintains that board members are not interested in hearing researchers' opinions. "They're just doing this as a political smokescreen," he says.

The Kansas case comes amid a wave of efforts by religious conservatives to limit the teaching of evolution nationwide. In October last year a school board in Dover, Pennsylvania, passed a policy requiring teachers to describe evolution as "not a fact". Other states, such as Alabama, are now revising their science standards to diminish the role of evolution.

McDonald says that Kansas Citizens for Science is planning a response to the May hearings. The board of education will decide on a final set of standards in June. ■

Viennese lab renovations stall as cash goes unspent

Quirin Schiermeier, Vienna

One of Europe's few institutes for the study of birds and animal behaviour is being hamstrung, some members of its board complain, as hundreds of thousands of euros meant for its refurbishment are going unspent.

Plans to bring the Konrad Lorenz Institute of Ethology in Vienna into the twenty-first century have been around for five years. But critics say that promised, and much-needed, investment in new staff, facilities and research has stagnated since 2002, when Dustin Penn, a behavioural biologist formerly at the University of Utah in Salt Lake City, took over as director.

Penn promised the Austrian Academy of Sciences, which runs the institute, that he would introduce cutting-edge genetic research, while building on traditional strengths such as ornithology.

But members of the institute's board say they have become increasingly concerned about the slow pace of change. *Nature* has seen a copy of the annual report, which shows that in the past two years Penn has returned around half of the institute's €1.3 million (US\$1.7 million) annual budget unspent. But *Nature* was unable to confirm these numbers with the

academy, which, even though its work is funded by the taxpayer, says these records are confidential.

Critics claim that Penn has spent most of his time managing a \$4-million project to investigate the genetic basis of human odour, funded by the US Defense Advanced Research Project Agency. They add that this project, carried out mainly in the United States and Britain, leaves Penn little time for the institute.

"If he is overstretched, the best solution would be for him to let someone else manage the institute," says board member John Dittami, head of the ethology department at the University of Vienna.

Penn denies that his workload is

unmanageable. "I spend a tremendous amount of my time running the Konrad Lorenz institute, supervising postdocs and negotiating with the academy," he says. "I am working overtime to get everything done."

But progress at the institute has stalled. Some of the draughty, low-ceilinged wooden buildings — a hangover from the Second World War barracks from which the institute was formed — do not meet modern research requirements. Some facilities are in desperate need of repair, including a leaking diving tank used to study fish. And although a small genetics lab has been installed, along with a temporary facility for housing mice, the planned expansion of genetics research has not yet come to fruition.

Penn blames unforeseen architectural problems and lack of permission from the Viennese planning authorities for the delay. "I have budgeted money for new equipment, but we couldn't buy everything that we planned simply because there's nowhere we could put it," he says.

The institute's board is set to meet this autumn to discuss the problems. But neither the board nor the academy would reveal to *Nature* exactly when an external review will take place. Georg Stingl, the academy secretary in charge of mathematics and natural sciences, says the evaluation might be "this or next year". ■

Additional reporting by Alison Abbott.



Cornered: critics say lab head Dustin Penn is overstretched.

Space agencies mull over mutual mission to Europa

Tony Reichhardt, Washington

A future probe to Jupiter's moon Europa and a space-based X-ray observatory should be built as joint ventures between NASA and the European Space Agency (ESA) over the next decade, researchers and agency officials say.

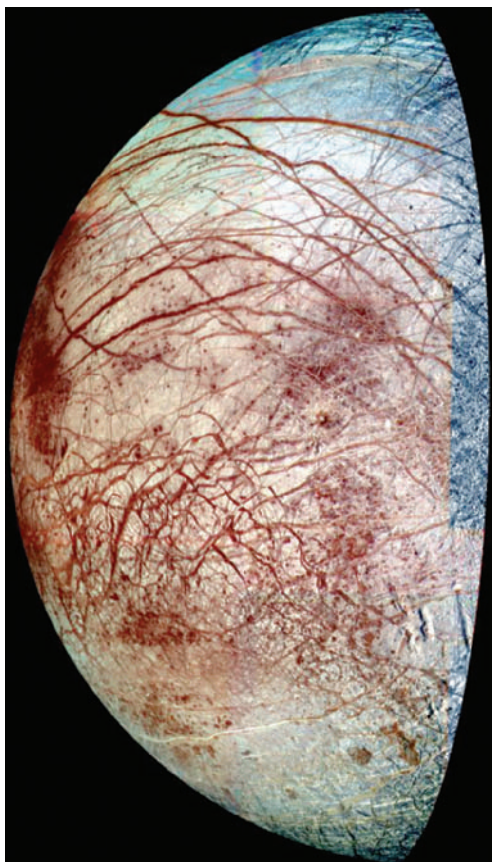
But both agencies will find it a challenge to amass cash for the projects, which will each cost between US\$1 billion and \$2 billion.

The joint missions were discussed earlier this month at a workshop near Washington that brought together representatives of the world's main space agencies to find where their research agendas overlapped.

Europa exploration and X-ray astronomy emerged as promising areas for cooperation, says Giovanni Bignami, director of the Laboratory of Space Astrophysics (CESR) in Toulouse, France, and chair of ESA's Space Science Advisory Committee.

NASA already has a mission in the design stage that would use a cluster of X-ray satellites to study dark matter and high-energy astronomical sources. ESA and Japan are considering a project with similar goals, called XEUS (for X-Ray Evolving Universe Spectroscopy). The Constellation-X Observatory, as the US mission is called, ranked highly in the most recent decadal review of priorities for US astronomy (see *Nature* 405, 381; 2000). XEUS is likely to figure prominently in a similar European plan to be released by Bignami's committee in May. Both missions are planned for launch in about ten years' time, and the two teams have already been collaborating informally in planning scientific objectives, says Harvey Tananbaum of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts.

Europa is a high priority for planetary astronomers owing to the liquid ocean thought to lie under its icy crust. US scientists want to keep Europa in NASA's sights following cancellation of the nuclear-powered Jupiter Icy Moons Orbiter (see *Nature* 433, 342; 2005). ESA has been funding a low-level study of technologies necessary to explore Europa, which could include a small probe sent to its surface. Alessandro Atzei of the agency's ESTEC science centre in the Netherlands says the one-kilogram package originally considered would be too small to have much scientific value. So ESA is looking at slightly bigger surface probes, along with



Ice lolly: Europe and the United States must find funds for the mission they are plotting to Jupiter's moon.

an orbiting radar to sound beneath the ice.

Whether a joint US–European mission would carry plutonium-powered batteries, as all previous NASA missions to the outer Solar System have done, is an open question. The joint Cassini mission to Saturn, which is considered a model for future cooperation, used them. ESA's current study is leaving open the possibility of using advanced solar arrays to generate power, as the European agency is not currently able to build plutonium batteries. But it would be natural for NASA to supply them on a joint mission.

All such details await a formal move by the agencies to set up a joint mission study — the workshop only got the ball rolling, says Bignami and others. Yet the willingness to cooperate is real, he says. And this time, following Europe's recent success in orbiting Mars and landing on Titan, the two space agencies will be more equal partners than ever before. In the past, says Ron Greeley, a planetary scientist at Arizona State University, Tempe, and long-time supporter of Europa exploration, the more experienced NASA tended to be the senior partner. Now, he says, "the balance has shifted".

Political aide sails into top job at maritime institute

Declan Butler, Paris

A scandal that prompted the resignation of the French finance minister, Hervé Gaymard, last month has ricocheted into the most unlikely of places — the national marine research agency, IFREMER.

Jean-Yves Perrot, who until the scandal broke was Gaymard's political adviser, was nominated by the agency's board on 22 March to take over as the agency's chief executive. Two days later the council of ministers confirmed his appointment.

The move follows a government decision earlier this month not to renew the mandate of the agency's current director-general, Jean-François Minster, a renowned oceanographer.

Perrot is an unknown face in French science. He lost his adviser's job when Gaymard was toppled in a controversy over rents. The French government paid the €14,000 (US\$18,000) monthly rent on a flat where the minister lived with his family, even though he owned other homes.

Several researchers have expressed outrage at Perrot's appointment, claiming that IFREMER is a victim of the fallout of the Gaymard scandal — especially as Minster's mandate at IFREMER had been viewed positively.

"I am profoundly shocked by the decision," says Pierre Papon, a socialist and former head of both IFREMER and the CNRS, the French national research agency, who is now a physicist at the Industrial Physics and Chemistry Higher Educational Institution in Paris.

IFREMER has traditionally been headed by scientists, whereas Perrot has had a solely political career. He is currently the mayor of Marly-le-Roi, a small town on the outskirts of Paris, and is a local representative in the greater Paris area for Jacques Chirac's ruling, neo-Gaullist UMP party.

"To nominate someone who has neither a scientific attachment, nor scientific or technical competence in the areas of activity of IFREMER, nor knowledge of public research in general, is a serious mistake and a serious political fault," says Papon. Perrot was not available for comment as *Nature* went to press.

For more news and analysis go to
news@nature.com
www.nature.com/news

'Libertarian' report condemns actions of fertility agency

London The future of Britain's Human Fertilisation and Embryology Authority, respected worldwide for providing balanced ethical oversight, has been called into doubt.

A cross-party group of politicians attacked the authority in a report issued on

24 March, claiming that it exerts excessive control over access to fertility treatment and is overly cautious in its approach to licensing embryo research. Critics have branded the report as too libertarian — for example, it says that couples undergoing *in vitro* fertilization should be allowed to choose the sex of their baby. And the report recommends that the authority's research and fertility-treatment arms be split into independent bodies.

But the Parliamentary Science and Technology Committee, which issued the report, was divided over the issues involved. Five of the eleven committee members declined to back it, saying that it was "light on ethics" and "goes too far in the direction of deregulation". The government will respond to the report in the next few months.

Switch in Indian patent law hits generic drug industry

New Delhi The Indian parliament has approved a bill to bring national patent rules in line with international norms. By requiring a product, rather than a process, to be patentable, the bill effectively wipes out the country's generic drug industry.

The opposition Bharatiya Janata Party and its allies walked out during the contentious debate in the lower house on 22 March. Last-minute measures were included to ease the transition and relieve some of the political tension. The bill was approved by the upper house the next day.

The World Trade Organization has been pressuring India for years to change its anomalous patent laws, which have allowed generics companies to grab a market of any new drug simply by changing a step in its patented synthetic pathway.

Space-elevator contest gives NASA a lift

Washington NASA is now in the contest business. Taking its cue from the Ansari X prize, which led to the first privately funded astronaut launch last year, the space agency last week announced its first Centennial Challenges. These are competitions meant to spur inventors to come up with breakthrough space technologies.

The awards will go for two advances related to space elevators — proposed systems for reaching orbit with long tethers anchored to a platform on Earth (see right). A US\$50,000 first prize will go to whoever develops the strongest tether material, with another \$50,000 to whoever can use a light source to push a robot 'climber' to the top of a 60-metre cable in less than three minutes.

NASA will put up a total of \$400,000 in 2005 and 2006 for the two contests, which have already



been announced by its partner in the challenges, the Spaceward Foundation in California.

AIDS advocacy groups say that the move will hit infected people in poor countries, who rely on the cheap generic drugs.

Nuclear soup cans put agency on the warpath

Washington The bizarre practice of keeping plutonium in soup cans, paint tins, plastic bags and other haphazard containers in US labs should no longer be tolerated, says the government agency charged with keeping an eye on safety at nuclear-weapons labs.

In a report published on 21 March, the Defense Nuclear Facility Safety Board criticizes the Department of Energy, which oversees the labs, for tolerating the practice, and calls for agency-wide, science-backed rules for storing plutonium.

Containers for radioactive leftovers in long-term storage are rigorously defined — storage must be in nested stainless-steel containers, welded shut and filled with inert gas. But there is no clear rule about how plutonium that is still being used in research should be stored. Aside from the obvious danger that lids could come off makeshift containers if they are dropped, the report notes that off-the-shelf packages could rapidly oxidize or corrode, and could allow radioactive gas to build up.

Biologists join together to speak the language of RNA

Washington What's in a name? A lot, according to molecular biologists who are striving to create a vocabulary especially to describe RNA molecules.

Announced on 21 March, the RNA Ontology Consortium is a network of RNA researchers who aim to develop a language to describe and catalogue RNA molecules.

Integrating databases of RNA sequences and three-dimensional structures will be a major focus of the five-year project, which is being funded by the US National Science Foundation to the tune of US\$500,000.

RNA is involved in a vast range of cellular functions, including conveying recipes for making proteins. Consortium head Neocles Leontis, a chemist at Bowling Green State University, Ohio, notes that proteins act as 'hardware' and are shared by many species, but the crucial differences lay in the 'software'. "We are beginning to see that RNA is that software," Leontis says.

Astronomer swaps one royal for another

London Astronomer Martin Rees is to be the next president of the London-based Royal Society. The University of Cambridge



Martin Rees is to take the reins at the Royal Society.

researcher, whose interests include γ -ray bursts and black-hole formation, will become the society's 59th president in November.

Rees, 62, will resign his current post as England's astronomer royal but retain his Cambridge professorship

and role as master of Trinity College.

He succeeds Robert May, a zoologist at the University of Oxford and Imperial College, London.

Asked whether the nomination was a surprise, Rees says it "depends on whether you believe what you read in the newspapers". He has been linked to the position several times in recent months by the British press. Rees adds that he is currently formulating the main themes of his presidency, which he will discuss in detail in November. "Until then I'll be doing a lot of listening and learning," he says.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The safety catch

The past year has seen a beleaguered Food and Drug Administration publicly denounced as unable to protect the US public. As the political pressure mounts, Meredith Wadman joins the the agency's hunt for a remedy to its ills.

Atelling scene unfolded in the lobby of a suburban Maryland hotel on 18 February. On the third and final day of a gruelling public conference, expert advisers to the US Food and Drug Administration (FDA) held in their hands the fates of three blockbuster painkillers, known as COX-2 inhibitors. To make their assessment, the experts had to balance the drugs' value in treating arthritis against mounting evidence that they cause heart attacks and strokes. Billions of dollars in revenues for drug companies — and the welfare of millions of patients — hung in the balance.

But when a mid-morning break was called and the 32 advisers wandered out for coffee, they were largely ignored by the throng of reporters at the event. Instead, the journalists clustered around a boyish-looking FDA scientist who was holding forth for a television

camera. They plied him with questions about his views on the issues before the panel, diligently jotting down his every response.

In the middle of the media scrum was David Graham, a career physician at the FDA. Graham came to public prominence last November at a US Senate hearing, when he pronounced his employer “incapable” of ensuring the safety of drugs after they were approved for sale in the United States.

His scathing testimony set alarm bells ringing on Capitol Hill and raised some fundamental questions about the FDA: is the agency scientifically, structurally or politically capable of ensuring the safety of some 10,000 pharmaceuticals now used by Americans? Or has the complexity, size and pace of the business, and the power of the drug manufacturers, exceeded the regulator's ability to cope?

Graham had told senators on 18 Novem-

ber that just one of the COX-2 inhibitors — Vioxx — had caused at least 26,000 deaths from heart attacks in the five years before Merck withdrew it from the market last September. An epidemiologist in the FDA's Office of Drug Safety, Graham had conducted a study using the massive database of Kaiser Permanente, a health-maintenance organization based in California. Last August, armed with its results, he warned his bosses that high doses of Vioxx significantly increased the risk of heart attack and sudden death. The agency did not act — and even approved the drug's use in children with rheumatoid arthritis in the weeks before Merck withdrew it.

The FDA's Office of New Drugs — where any new medicine must pass muster before it can be sold — approved Vioxx in 1999. It was heavily promoted by Merck as an alternative to standard arthritis medications because it didn't cause stomach bleeding, and it soon became a huge success, with worldwide sales of \$2.5 billion in 2003 alone. Then, last September, a Merck-sponsored trial examining whether the pill might prevent precancerous colon tumours found that Vioxx doubled the risk of heart attacks and strokes in patients using it for more than 18 months.

Critics — including plaintiffs in the raft of lawsuits against Merck that swiftly followed the drug's withdrawal — complained that as early as 1999, studies had flagged Vioxx's potential for causing heart attacks and strokes. They implied that the FDA should

R. JAMIE/GAMMA

have identified the problem and pulled the medicine from the market.

In his Senate testimony, Graham endorsed that point of view. He added that Vioxx was probably far from a one-off event. It was bound to repeat itself, he said, because the FDA's organizational structure and corporate culture were biased towards approval of new drugs. Safety monitoring of drugs already out there, he said, took second place.

Graham further complained that the reviewers at the Office of New Drugs who approve therapeutics in the first place have a vested interest in the drugs' success, and are liable to ignore or overrule post-market safety concerns raised by staff scientists in the smaller Office of Drug Safety. Asked to identify other established drugs about which he had serious safety reservations, Graham named five, including Pfizer's Bextra, another COX-2 inhibitor. "The FDA as currently configured," he concluded, "is incapable of protecting America against another Vioxx. We are virtually defenceless."

Rapid response

The initial congressional response was swift and emphatic. Senator Chuck Grassley (Republican, Iowa), chairman of the Senate Committee on Finance where Graham testified, chided the FDA for ignoring danger signals and failing to heed the warnings of its own scientists. Senator Michael Enzi (Republican, Wyoming), who chairs another Senate committee with direct jurisdiction over the FDA, scolded its officials at a hearing earlier this month saying that "doing nothing to address the current controversies is not an option". Grassley is shortly expected to introduce legislation, together with Senator Christopher Dodd (Democrat, Connecticut), that would boost the power and autonomy of the Office of Drug Safety.

Yet top FDA officials say that the agency's performance remains strong. "The safety of the drug supply right now is better than it has ever been," says Janet Woodcock, the FDA's acting deputy commissioner for operations, who from 1994 to 2004 ran the agency's Center for Drug Evaluation and Research (CDER), which houses both the Office of New Drugs and the Office of Drug Safety.

Woodcock argues that some agency critics fail to understand that risk and benefit are inextricably linked for any drug — and that calls to assess one without the other betray this basic misunderstanding. The point, she says, isn't that some drugs with risky side effects shouldn't be approved, but that their risks and benefits need to be carefully weighed both by the regulatory authorities who first approve them and by the physicians and patients who ultimately use them.

And despite the wave of bad publicity that

shrouded the FDA last winter, its public image remains fairly healthy. For decades, polls have suggested that its public approval hovers at the heady level — for a government agency — of about 75%. And a February survey of 1,200 people by the Kaiser Family Foundation found that 77% thought the agency was doing a reasonable job of ensuring drug safety.

Nonetheless, by almost any measure, the past year has been a rocky one for the FDA. In March 2004, Mark McClellan decamped after just 17 months from a commissioner's position that had sat vacant for nearly two years prior to his arrival. Soon after that, the agency came under fire for suppressing the

"The FDA as currently configured is incapable of protecting America against another Vioxx. We are virtually defenceless."

— David Graham

report of a staff scientist who cautioned that popular 'SSRI' antidepressants could cause suicidal tendencies in young people. And in early October, just a few days after Merck withdrew Vioxx, the FDA's competence in ensuring the safety of vaccines destined for the US market was called into question when British authorities abruptly closed a flu vaccine plant in Liverpool owned by US company Chiron. This halved the United States' supply of vaccine for the coming winter and provoked a public uproar.

David Kessler, who ran the agency from 1990 to 1997 and is now dean of the medical school at the University of California, San Francisco, thinks that the cumulative impact of events has been considerable. "It certainly

began to shake confidence, not only in the public but within the medical community," Kessler says. "For the first time, I have physician colleagues asking me if they can believe what the FDA is saying."

Full stretch

Some of the 1,600 scientists and physicians who work at the CDER privately admit that they are stretched to fulfil the agency's mission. With an average annual salary of \$128,000, many are forsaking more lucrative careers in the private sector — often because they find FDA work satisfying and valuable.

"I tell my kids I have one of the most interesting jobs in the world," says Rachel Behrman, a physician who began working as a drug reviewer in 1989 and is now deputy director of the CDER's Office of Medical Policy. "And I go to bed every night thinking I've done something important."

But the current crises, combined with the fact that 30% of the FDA's work force is eligible for retirement over the next five years, has agency-watchers worried that retention of high-quality staff is going to become a tough challenge. "If we continue to pound and beat up on the agency, we are going to lose very good people," says Kessler.

Worries about the pressure on FDA staff surfaced in 2002, when the Government Accountability Office assessed the impact of a 1992 law that introduced industry fees to pay for speedier drug reviews. It found that

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS



Balancing act: Janet Woodcock argues that the key to drug regulation is weighing up risks and benefits.

the law had boosted reviewer workloads, cut their training and resulted in annual attrition rates among FDA scientists of about 10% — roughly twice the level found at the National Institutes of Health or the Centers for Disease Control and Prevention.

And a survey of 400 FDA scientists conducted for another 2002 government study — but not published until December 2004, when it was accessed by the Union of Concerned Scientists under the Freedom of Information Act — found a group of people seriously concerned about the agency's ability to protect the public from dangerous side effects in approved drugs. In that survey, 19% said that they were "not confident at all" in the FDA's ability to monitor the safety of drugs once they are on the market; only 6% were "completely confident" in that ability. And 18% said that they had been pressured to approve new drugs despite reservations about their safety, efficacy or quality.

Testing times

The COX-2 controversy has left some observers asking whether the CDER, with an annual budget of about \$500 million, has the resources to police a prescription drugs market that is now worth more than \$160 billion annually in the United States. Some also argue that the FDA lacks the legal authority it needs to monitor established drugs.

Before a drug is approved, the onus rests on the manufacturer to demonstrate its safety and efficacy. Once the FDA approves a medicine for marketing, that balance shifts. The agency can ask — but not require — companies to pay for and conduct post-market safety studies; and it cannot limit the use of a drug to particular medical subspecialties, as it can with medical devices. Nor does it have explicit

authority to beef up warnings on drug labels. Instead, it negotiates label changes in often-protracted discussions with manufacturers. In the case of Vioxx, a change to the label to include a "precaution" about the risk of heart attack at high doses took 18 months for the agency to negotiate with Merck.

That kind of outcome begs for additional legal muscle for the FDA, says William Schultz, a Washington lawyer who was the agency's deputy commissioner for policy in the mid-1990s. "It's the post-market piece that really needs the attention," he says. "Patients have got to realize how little we know when a drug goes on the market."

There are other thorny problems in ensuring post-market safety. The FDA, for example, doesn't set out systematically to monitor dangerous side effects. Instead, it relies on a reactive, 'passive' reporting system in which doctors report possible cases of side effects only when it occurs to them that a particular ailment may be the reaction to a drug. The system, by its nature, will seldom detect dangerous side effects that are already common in the population, such as the heart attacks and strokes prompted by the COX-2 inhibitors (see page 557). It is thought that the 400,000-odd reports that the FDA gets in this way each year represent only a small fraction of actual adverse events.

Then there is the fact that the 1992 law that instituted industry-paid fees did so on the condition that the FDA met tight drug-review timelines and otherwise boosted performance standards. To achieve this — and so keep hundreds of millions of dollars in industry fees flowing into the agency — the drug centre was forced to raid other

budgets. The upshot: in President Bush's proposed 2006 budget, new-drug review consumes roughly 80% of the CDER's budget; only about 6% of it is designated for post-market surveillance.

The 1992 law "sapped resources from other very needed areas, away from drug safety, away from compliance", says Kessler. Post-market monitoring "absolutely needs more resources. You can't just have resources go into new drug review", he adds.

Independent thought

In at least some quarters on Capitol Hill, that message is being heard. The bill being drawn up by Grassley and Dodd would create a fully funded Center for Drug Safety independent of the CDER with authority to demand label changes from drugmakers. "It doesn't make sense to have the office that reviews the safety of drugs under the thumb of the office that puts the drugs on the market in the first place," Grassley told the Consumer Federation of America last month.

Acting FDA commissioner Lester Crawford, who has been nominated by President Bush for the permanent post, told senators at his nomination hearing on 17 March that he is "open to discussing" an independent office of drug safety. Crawford has also announced that a new board for overseeing drug safety — an advisory board of mainly FDA employees — will publicize worrisome side effects more quickly than has happened in the past. Critics immediately assailed the board as toothless, because it will lack the power to require label changes or to pull drugs from the market. But Crawford says that it will herald a new era of public openness at the agency.

Still, some agency-watchers fear that the relatively low profile of Crawford's nomination and confirmation process, together with the reluctance of the conservative Congress to antagonize its allies in industry, suggest that even the tens of thousands of deaths attributed to Vioxx may be insufficient to initiate any real strengthening of the FDA.

The agency's history demonstrates that "major changes come after disasters", says Schultz. And the Vioxx case may not rise to that threshold in the public mind. At the end of the three-day advisory panel meeting back in February, the experts voted narrowly to allow Vioxx back onto the market, subject to careful constraints. They also voted to give patients continued access to Pfizer's two COX-2 inhibitors, Celebrex and Bextra.

Graham, for one, was not impressed. "Despite the biggest drug-safety catastrophe in the history of the United States," he noted with his trademark earnestness, "people are heavily invested in the status quo."

Meredith Wadman is a freelance writer based in Washington DC.

"For the first time, I have physician colleagues asking me if they can believe what the FDA is saying."

— David Kessler

Chasing shadows

Why do current safety systems sometimes fail to notice that approved drugs are causing serious adverse effects? And can surveillance methods be improved? Simon Frantz investigates.



It's 3 a.m. in a typical US emergency room, and doctors are bustling around a patient in his mid-sixties, one of several admitted that night with severe chest pain. He's anxious, sweating and struggling to breathe — classic symptoms of a heart attack.

By the following evening, he has stabilized, thanks to the prompt administration of streptokinase, which dissolved the clot that was starving his heart muscle of oxygen. It's time to examine factors such as obesity, smoking and blood cholesterol that might have contributed to the attack, and advise the patient about lifestyle changes that will help minimize the chances of a recurrence.

An estimated one million Americans are admitted to hospital under similar circumstances each year; about a third of them die. This is an all-too-common emergency, with a host of everyday causes.

No one thinks to consider whether the patient's attack might have been triggered by the pills, prescribed to ease the pain of arthritis, sitting on his bedside table back at home.

At least that was the case before 30 September 2004, when the pharmaceutical giant Merck withdrew its painkiller Vioxx from the market after it was linked to an increased risk of heart problems. The news

was met with confusion, anxiety and outrage. How could the Food and Drug Administration (FDA), which is supposed to protect the US population from unsafe medicines, and which approved Vioxx for sale in May 1999, not have noticed the dangers?

Hidden dangers

Critics argue that the FDA needs to re-examine its priorities in light of the Vioxx scare (see page 554). But whatever the outcome of that debate, the drug's withdrawal has revealed how adverse effects such as heart attacks, which already occur commonly in the general population, can slip under the radar of current drug-safety surveillance. And in the wake of the Vioxx controversy, experts are examining ways in which methods of detecting such effects could be improved.

The problem with Vioxx was never going to emerge in the phase III clinical trials that are used to judge whether a drug should be put on the market. The heart attacks occurred in less than 2% of patients who had taken the drug for many months¹, whereas phase III trials are typically carried out on a few thousand patients over a

timescale of several weeks. Only when hundreds of thousands of patients had taken Vioxx for an extended period would any problems become apparent. Such events can only feasibly be investigated after a drug is in general use.

Current post-marketing surveillance systems can easily detect adverse events that are unexpected and rare. For instance, the multiple sclerosis treatment Tysabri was withdrawn from the US market in February, about four months after it was given expedited approval, when two people contracted a rare brain disorder called progressive multifocal leucoencephalopathy, or PML. "PML is a very rare event, so this was like a red flag waving," says Steven Galson, acting director of the FDA's Center for Drug Evaluation and Research.

Invisible enemy

The difficulty with common conditions such as heart attacks is that cases that might be triggered by prescription drugs readily get lost in the noise. That's especially true for patients with ailments such as arthritis, who tend to be older, and therefore more susceptible to heart disease. Conventional post-marketing surveillance schemes, which mainly rely on doctors informing drug companies of adverse events associated with their products, and companies in turn informing regulators, are almost useless in this regard.

"Spontaneous reporting systems cannot distinguish whether adverse events such as

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Bitter pill: health problems linked to Vioxx exposed the frailties of post-market drug safety surveillance.

heart attacks are the result of a drug or whether it would happen in that community anyway," says Alasdair Breckenridge, who chairs the UK Medicines and Healthcare products Regulatory Agency.

In such cases, the burden falls on researchers in the field of pharmacoepidemiology, who study clinical databases containing millions of patient records, documenting information on medicines prescribed, hospitalizations and deaths. Taking their cues from potential problems highlighted in clinical trials, case reports or lab experiments, these researchers use statistical tools to study large patient populations over extended periods.

"We look through the databases to find an exposed group and a non-exposed group that are quite similar except that one group gets Vioxx, for example, and one group doesn't," explains Susan Jick, an epidemiologist at the Boston University School of Public Health, and co-director of the Boston Collaborative Drug Surveillance Program, based in Lexington, Massachusetts. "Or you do a case-control study where you look at similar people who have had, say, a heart attack, and those who haven't, and you see whether they have taken Vioxx or not."

This sounds simple. But in practice, it is a painstaking and labour-intensive process. The problem is particularly acute in cases in which it is difficult to determine whether an adverse event is a consequence of the underlying disease, or the medication — for example, when studying patients with heart attacks who are taking drugs to control their blood pressure. But even in less ambiguous situations, it's tough to sort patients into groups that can fairly be compared to one another.

Class divides

In the case of Vioxx, which belongs to a class of drugs called COX-2 inhibitors, concerns about an elevated risk of cardiovascular disease were first raised in 2001 by a reanalysis of a study comparing patients on drugs from this class with those on an older painkiller called naproxen². At the time, Merck suggested this was because of the cardioprotective effect of naproxen, rather than adverse effects of Vioxx.

Researchers led by epidemiologist Wayne Ray of the Vanderbilt University School of Medicine in Nashville suspected otherwise. Their initial study, which examined patients on naproxen and other drugs of the same class, known as non-steroidal anti-inflammatory agents, found no evidence of a cardioprotective effect³. But a direct examination of the risks posed by Vioxx, conducted by a collaboration including Ray's team, and led by FDA researcher David Graham, wasn't published until 2005.



Wayne Ray's team demonstrated that long-term use of Vioxx increases the risk of a heart attack.

The study⁴ suggested that up to 140,000 excess cases of serious coronary heart disease — resulting, Graham says, in at least 26,000 deaths — could have occurred in the United States while Vioxx was on the market. But this required painstaking effort to decide which patients to include in the comparisons, examining such variables as existing cardiovascular disease and arthritis. The project drew on the skills of a cardiologist, a rheumatologist and a statistician. "You have to have all these experts who can bring their knowledge to bear," says Ray. "It might take a few months to do the computer runs, but it takes many, many months of careful thought going into it beforehand."

Such difficulties are compounded by a dearth of support for pharmacoepidemiological studies. "The frustration is that we have the technology, but there is very little funding for carrying out these studies," says Brian Strom, an epidemiologist at the University of Pennsylvania School of Medicine in Philadelphia.

US government agencies don't provide significant funding for such studies. The FDA's Office of Drug Safety, which is responsible for monitoring and assessing the safety of existing drugs, has extremely limited funding for epidemiological studies of its own. A small programme at the Agency for Healthcare Research and Quality allocates just \$5 million for drug-safety surveillance studies — a fraction of the roughly \$450 million it costs on average to do the clinical trials on a single drug before it can be brought to market.

Aside from funding for more studies of

drug safety, the main need is to develop clinical databases that pharmacoepidemiologists can readily query to test their hunches about adverse drug events. These need to incorporate full records of patients' medical history, including any hospitalizations and the medicines they were prescribed.

The largest such database is the UK General Practice Research Database, which contains over 2.5 million patient records. It is a product of Britain's nationalized healthcare system, in which family doctors are patients' first point of contact. This database has proved successful, for instance, in confirming fears that antidepressants might trigger suicidal behaviour in some patients⁵. But ideally researchers want to extend studies to the United States, where drugs are typically approved earlier, and are taken up more quickly by larger numbers of patients.

Separate sources

Here, the problem is that such data are collected by individual Health Maintenance Organizations (HMOs), such as the California-based Kaiser Permanente, whose database was used for Graham's Vioxx study. Integrating these databases, and ensuring that the data they contain have common definitions and formats, would create a powerful resource for observational studies — but it is a major challenge. "Integration is probably the biggest stumbling block," says Donald Berry, who chairs the Department of Biostatistics and Applied Mathematics at the M. D. Anderson Cancer Center in Houston, Texas.

Nevertheless, some attempts at database integration are now getting off the ground. Eric Larson, director of the Center for Health Studies in Seattle, part of the non-profit Group Health Cooperative, is developing the Coordinated Clinical Studies Network. Involving 13 HMOs around the United States, the aim is to create an integrated data 'warehouse' containing clinical information on up to 18 million people, initially focusing on cardiovascular disease. The three-year project began in November 2004 and has been funded to the tune of \$2.7 million by the National Institutes of Health.

Even with the funding hurdle cleared, other problems might hamper this ambitious initiative. The biggest difficulty is the lack of will to carry out this project, says Larson. For people working in private sector HMOs, devoting up to three years to setting up a network, rather than actually using it, is a daunting prospect. "This is a labour of public service," says Larson.

Simon Frantz is news editor for *Nature Reviews Drug Discovery*.

1. Bresalier, R. S. *et al.* *N. Engl. J. Med.* **352**, 1092–1102 (2005).
2. Mukherjee, D., Nissen, S. E. & Topol, E. J. *J. Am. Med. Assoc.* **286**, 954–959 (2001).
3. Ray, W. A., Stein, C. M., Hall, K., Daugherty, J. R. & Griffin, M. R. *Lancet* **359**, 118–123 (2002).
4. Graham, D. J. *et al.* *Lancet* **365**, 475–481 (2005).
5. Jick, H., Kaye, J. A. & Jick, S. S. *J. Am. Med. Assoc.* **292**, 338–343 (2004).

Time to speak up for climate-change science

If debate is left to greens and sceptics, people think the evidence is equal on each side.

Sir— The science of climate change is under attack — an attack that is coordinated, well-funded and given constant play in the media. The stronger the scientific consensus on climate change becomes, the more the media suggest that the science is uncertain.

Lobby groups such as the Scientific Alliance and the International Policy Network have sprung up to offer climate ‘sceptics’ to the media. In the United Kingdom, the BBC appears incapable of hosting a discussion on climate change without letting one of these people claim that it is not happening.

The impression created in the public

mind is that climate scientists are deeply divided, and action to reduce greenhouse-gas emissions would be premature.

Yet the consensus among climatologists, glaciologists and atmospheric physicists — that anthropogenic climate change is a reality — is as robust as is likely to be found on any scientific issue.

As environmental campaigners, we would like to ask climate scientists everywhere: why are we being left to carry the can?

We’re not asking you to become campaigners or to compromise your independence. But we wish you would defend your profession as any other

professionals would. This includes training people in media relations, sending eminent delegations to meet editors and senior journalists, writing letters to the papers to correct misleading articles and seeking every opportunity to put the record straight.

Isn’t it time you started fighting for your science?

George Monbiot

82 Fairacres Road, Oxford OX4 1TG

Other signatories of this letter:

Mark Lynas 30 Ullgar Road, Oxford OX2 8AZ

George Marshall Climate Outreach Information Network

Tony Juniper Friends of the Earth UK

Stephen Tindale Greenpeace UK

We Africans must take our future in our hands

Sir— Your Editorial “Africa 2005” (*Nature* **433**, 669; 2005) stresses that it needs to be Africans, not donors, who decide how to spend the billions of dollars Africa is to receive in aid. But money alone is not what the continent needs for sustainable development. African countries have received billions of dollars in aid and loans during the past half century, and the result is what we have today.

The continent needs good management, relevant projects and a boost to science and technology. Regional economic integration combined with increased exports of our products and competitive investments are enough to sustain economical development.

Many people in Africa are aware of the real problems and their solutions. In 2000, when the Department of Education in Rwanda negotiated scholarships with a non-governmental organization for 100 students to study in the United States, the minister of education told us that the country needs scientists and engineers.

Before coming to the United States to study computer science under this scheme, I had always thought that Africa’s under-development and poverty were mainly caused by Western nations through slavery, colonization and imperialism. This concept stemmed from the information I had at the time. But even if it contains some truth, I have come to realize that the current situation depends on what we Africans are doing in our continent and on our attitudes towards international exchanges.

To help poor countries is not bad. But Western nations can’t help us to become as

developed as they are. I share a vision with *Nature* and with John Mugabe, the scientific adviser to the New Partnership for Africa’s Development: the future of Africa is in the hands of Africans.

Justin Hategekikimana

8228 Ohio River Boulevard, Apt 45, Pittsburgh, Pennsylvania 15202, USA

Driving passion brought rare bird to the masses

Sir— Your reprinted item from 1955 concerning an expedition to Sierra Leone to find *Picathartes gymnocephala* (“50 years ago” *Nature* **433**, 23; 2005) did not mention that the team was led by a recently appointed member of BBC staff called David Attenborough.

The search formed the central plank of the *Zoo Quest* series, which captured the imagination of the British television-watching public in the 1950s and was the direct forerunner of classic programmes such as *Life on Earth*, which was broadcast to millions around the world.

In his autobiography *Life on Air* (BBC Books, London, 2002), Attenborough tells of a bus driver pulling up next to the open-top car in which he was travelling. “Ere, Dave,” called the bus driver, “are you going to find that Picafartes gymno-blinking-cephala or aren’t you?”

He did — and as a result, through increased public awareness and government funding, we have all discovered much more since then.

Rupert C. Marshall*, **Helen E. Nice†**

*Institute of Zoology, Gothenburg University, Medicinaregatan 18, 41390 Gothenburg, Sweden

†Marine Monitoring Unit, Sydney Water, 51 Hermitage Road, West Ryde, 2114 NSW, Australia

Authors say that they prefer online submission

Sir— John Moore, in Correspondence (*Nature* **433**, 800; 2005) may be right that online submission makes authors produce their own manuscripts and figures, but the evidence is that they overwhelmingly prefer this to other means.

My company is conducting a survey, commissioned by the Association of Learned and Professional Society Publishers, of recently published authors, using a sample provided by ISI. Of the 442 responses received to date, more than 80% say that online is easier than hard-copy submission. A similar number prefer online submission to any other method, and 36% say that the lack of an online system would influence them against submitting to a journal.

Respondents named a number of benefits including faster responses, shorter refereeing times, greater transparency and convenience. They did perceive areas for improvement, particularly in ease of use, system performance and flexibility, but overall they think the benefits greatly outweigh the problems.

Further, the editors and publishers we have surveyed say that the introduction of online systems leads to an increase in submissions from less-developed regions.

The full report is due to be published at the end of April.

Mark Ware

Mark Ware Consulting Ltd, 14 Hyland Grove, Westbury-on-Trym, Bristol BS9 3NR

Nature surveys all its authors immediately after publication to ask about their experience, in order to improve our services. Please contact authors@nature.com if you have any comments.

Confronting the human dilemma

How can ecosystems provide sustainable services to benefit society?

Harold Mooney, Angela Cropper and Walter Reid

Four years in the making, the Millennium Ecosystem Assessment (see *Nature* **417**, 112–113; 2002) is released this week (starting 30 March). This gigantic endeavour explores the link between human well-being, the status of ecosystems and their sustainable use.

What has this assessment taught us about developing our planet, and will it, or should it, be continued? To answer the first part of this question, the assessment is an invaluable record of where we stand now, and why. But for it to be useful, the answer to the second part of the question must be 'yes'. We need to take a consistent approach to measuring the status and trends of the world's ecosystems. To take one example, the Convention on Biological Diversity has set the target of reducing the rate of global loss of biodiversity by 2010. But the data to evaluate whether this goal is being met are not readily available, as biological diversity is more than just an enumeration of species present or absent — it includes parameters such as the populations of species and the ecosystems in which they reside. In addition, biodiversity is just one of the many aspects of change in ecosystems and their related functions assessed in the Millennium Ecosystem Assessment. Only from a periodic audit of the state of our natural resource base can we determine if we are indeed approaching sustainability.

At present, there are no formal plans to repeat the Millennium Assessment. There should be, and we hope that the informal discussions among the current sponsors will bear fruit from the seeds sown by the many smaller, ongoing sub-global assessments that were stimulated by the assessment.

Achievements and goals

Although there has been a steady increase in many indicators of human well-being in many parts of the world — such as an increase in personal wealth, a longer life-span and access to plentiful and inexpensive food — these benefits have not been universally distributed. There are still more than one billion people surviving on less than one dollar a day and nearly that many are undernourished. About 1.1 billion people lack access to a basic water supply and more than 2.6 billion lack access to basic sanitation. It is this disparity that has driven the United Nations to set the goals of halving the proportion of people living on less than one dollar a day, reducing the proportion of



Sustaining ecosystems does more than just aid conservation, it saves resources for human use.

people suffering from hunger and halving the proportion of people with no sustainable access to safe drinking water and basic sanitation. To achieve these goals by 2015, it is accepted that nations have to achieve sustainable development and reverse the losses of environmental resources.

The Millennium Ecosystem Assessment took a new pathway of evaluating the status of the Earth's human support systems. Rather than the standard environmental audit, the new assessment places audits of numbers of organisms and so on into the context of how ecosystem changes have affected human well-being, and how they may do so in the foreseeable future. It had to find a link between the status of biotic systems and the status of individuals in various societies in the world to estimate the capacity of ecosystems to provide services that benefit society. Many of these links are obvious, but others have not been appreciated, nor have all these linkages been quantified. In essence, we had to make a large leap from the current styles of evaluations of status and trends in ecosystems to an entirely different approach — an ecosystems services database related to how ecosystems and societies operate, and how they interrelate.

Current status of ecosystems

Human societies have made marked progress in increasing provisioning services, such as crops and livestock, to meet the demand of a growing population (see overleaf). Food is

more abundant and cheaper than in the past. Despite these dramatic accomplishments, there are still more than 850 million undernourished people, and some advances in production are at the cost of other services essential for human well-being, such as ocean fisheries, wood for fuel, genetic resources and — perhaps the most important — fresh water. It is the poor in many nations that are most directly dependent on services from ecosystems, and the degradation of these systems can exacerbate their poverty. Millions of people face the reality of the declining availability of cheap protein from local fisheries, inadequate water for sanitation or live on degraded landscapes.

There are a number of issues that cloud the goal of sustaining a high level of provisioning services. The use of fertilizer in agriculture has greatly increased to meet food demand, but at the cost of polluting off-site unmanaged ecosystems, such as groundwater, rivers and coastal fisheries. In many regions, water for irrigation is being pumped from groundwater and in some cases from fossil sources. Rivers are dammed and diverted for irrigation, altering ecosystems that depend on this water — causing the loss of many of the services they provided.

Further, we are diminishing crucial 'regulating' services responsible for climate, erosion, air- and water-quality control, as well as for the regulation of pests and natural hazards. We are losing these services due to massive land-surface conversion, atmosphere alteration, eutrophication, overharvesting and the impact of invasive species. The Millennium Assessment concluded that 60% of the ecosystem services evaluated were either being degraded or being used unsustainably.

As an example, cultivated systems (areas where at least 30% of the landscape is in croplands, confined livestock production or freshwater aquaculture) now cover a quarter of the Earth's surface, partly by conversion of temperate grasslands, Mediterranean-climate forests and many tropical ecosystem types. Forests have essentially disappeared from 25 countries, with 9.4 million hectares being lost annually from the Earth's surface. Historically important fisheries have collapsed or are overfished, one third of the mangrove forests for which there are historical data have been lost, as have 20% of the coral reefs, with a further 20% degraded. Nearly 40% of the rivers of the world have been fragmented. Species and populations of species are being lost at unprecedented rates, while at the same time the global biota is becoming homogenized owing to the introductions of alien

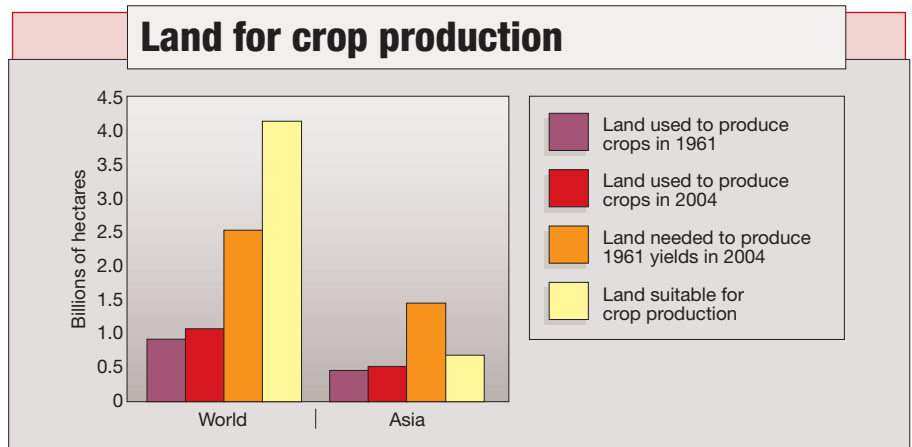
species to new regions. These examples represent major losses of pieces of the biosphere machinery, which have a serious impact on the delivery of ecosystem-regulating services — impacts such as greater prevalence of infectious diseases in disrupted ecosystems, adverse effects on local climates by ecosystem modification, and the loss of flood protection (as in the recent tsunami in Indonesia).

What we can do

The drivers of change in ecosystems and their services will continue in direction and intensity. So how can these trends be reversed to achieve sustainability and to relieve the negative impacts of the loss of services to society, particularly to the disadvantaged? New pathways and approaches can and must be taken. But these are major initiatives, which will mean profound changes in the way global society operates. As learned in the Millennium Assessment, favourable responses need to take place at all levels, from the local to the global. Global mechanisms do not necessarily solve local problems, yet are an important part of the overall solution. At the same time, local players and solutions can feed into regional and global approaches. The players at these different levels address different decision-makers, who can collectively put in place the major changes that are needed for ecosystem sustainability.

The Millennium Assessment examines the merits of options for mechanisms and policies, to accomplish the goal of maintaining and enhancing the delivery of ecosystem services to society. Some of these require major reorganization in the way we do business. At present, our organizational structures address separately the issues of a single resource, such as agriculture, fisheries or the environment. There is little interaction within and between each issue, and much less again with trade and the treasury bodies. The lesson of the Millennium Assessment is that all these resource issues are interrelated: action on one issue has consequences for another. It is crucial to address how to minimize the trade-offs (biodiversity or clean water for agricultural yield), either on-site or by managing landscapes. One important example of how this process can work is the EU system of directives for nitrate accounting on landscapes.

Some institutional innovations are moving towards more integrated views of issues and responses to them. For example, Britain has a government department for Environment, Food and Rural Affairs. These are all closely interrelated domains, but in other countries are often handled by competing agencies. Elsewhere, interagency groups are evolving to address central issues such as climate change, but their effectiveness is hampered by competitiveness and politics. We need new kinds of institutions in better



Cropland expansion versus intensification

A key trade-off in cultivated systems is between increasing the amount of cropland needed to meet growing food needs versus increasing the productivity of each hectare of cropland. The 'land-sparing' impact of modern farming practices has mainly been achieved by yield increases from use of crop monocultures with improved crop varieties, fertilizer inputs and irrigation. For example, if yields of the six major crop groups that are cultivated on 80% of the total cultivated land area had remained at 1961 yield levels, it would require an additional 1.4 billion hectares of land in 2004 — more than double the amount currently used (see graph). This represents 34% of total land area suitable for crop cultivation, and would have required conversion of large areas of uncultivated land that support rain forests, grassland savannahs and wetlands. In Asia alone, it would require an additional 600 million

hectares, which represents 25% more land area than is suitable for cultivation on this continent. Asia would now be heavily dependent on food imports if crop yields had remained at 1961 yield levels. Although this increase in productivity has saved some land from conversion, it has resulted in greater impact on other services through water withdrawals, excessive nutrient loads and pesticide use.

The key ecological question is therefore whether environmental services — other than food production at regional and global scales — would be enhanced by focusing food production on less land under intensive management with high yields, versus expanding cultivated area in lower-yielding systems using farming practices that preserve environmental services at the field and local levels. Few studies have addressed this issue using sound, ecological analytical methods.

positions to achieve sustainability of ecosystems that provide for human well-being.

We must also try to improve the economics. Although provisioning services are enmeshed in the local (and increasingly global) marketplace, regulating services are not. We must accelerate our ability to value ecosystem-regulating services at the national level, as well as the ecosystem services that provide crucial cultural amenities, and ensure that these values are considered in decision-making.

Some progress is being made. Costa Rica has established a system of conservation payments, under which contracts are brokered between international and domestic 'buyers' and local 'sellers' of sequestered carbon, biodiversity, watershed services and scenic beauty. On a global scale, the Ecosystem Marketplace consortium is beginning to track transactions, pricing trends and buyers' requests on the carbon, water and biodiversity markets. It is predicted that the global carbon market will reach US\$44 billion by 2010.

We need to eliminate the subsidies that promote the excessive use of ecosystem services and evaluate more carefully the

trade incentives that damage ecosystem services. We must work harder to educate the public on the strong links between sustainable ecosystems and the lives of humans. The role of new technologies in more efficient use of natural resources is crucial and needs more incentives.

There is plenty that can and needs to be done to deal with the crisis that has already enveloped us. The path is open for scientists to quantify, to a much greater extent, the way in which the operation of ecosystems is directly linked to human well-being, and hence model the course of human activities on future outcomes of the delivery of these services. The Millennium Assessment is certainly providing a strong stimulus for such studies. ■

Harold Mooney is in the Department of Biological Sciences, Stanford University, California 94305, USA. Angela Cropper is at the Cropper Foundation, Port of Spain, Trinidad and Tobago. Walter Reid is the Director of the Millennium Ecosystem Assessment, 4225 Glen Ave., California 94611, USA.

♦ www.millenniumassessment.org

Acknowledgements.

We thank the scientists, reviewers and members of the review board who provided input to the Millennium Ecosystem Assessment, and the sponsors of this work.

Words of wisdom

A scholarly selection of scientific soundbites and sayings.

Oxford Dictionary of Scientific Quotations

edited by W. F. Bynum & Roy Porter
Oxford University Press: 2005. 730 pp.
£30, \$50

Steven Shapin

In the popular mind, scientists are supposed to be people of few words, fewer still memorable. Most sayings by scientists that stick in the mind do so because at some point we have been made to memorize them, or at least some form of them. This is true for scientific laws ("To every action there is always an opposite and equal reaction"), principles ("Ontogeny recapitulates phylogeny") and more general sentiments ("Nature makes no jumps"). Modern scientists are not strong on rhetoric, but they are not supposed to be.

The motto of the Royal Society of London — "*Nullius in verba*" — is properly translated as "On no man's word", but some commentators construe it to mean something like "Don't use fancy language". The society's founders in the seventeenth century argued strongly against the sorts of stick-in-the-mind figurative speech in which their literary contemporaries trafficked. In the twentieth century, literary critic I. A. Richards pointed out: "We believe a scientist because he can substantiate his remarks, not because he is eloquent and forcible in his enunciation."

So the editors of the *Oxford Dictionary of Scientific Quotations* were well advised to assemble not just notable scientific sayings by scientists, but also comments on science, scientific objects and scientific practitioners made by non-scientists, and remarks by scientists on all sorts of things outside their specialized expertise. That allowed them to include W. H. Auden's comment "When I find myself in the company of scientists, I feel like a shabby curate who has strayed by mistake into a drawing room full of dukes," and Oscar Wilde's dismissive quip about the splendours of Niagara Falls: "It would be more impressive if it flowed the other way."

Nor were the editors much bothered about what properly counts as science. They trawl ancient philosophy (Aristotle's

"All men by nature desire to know"), proverbial lore, both traditional and modern ("Nature is by nature perverse" and "Garbage in, garbage out"), sociology (Émile Durkheim's "Consider social facts as things") and anthropology (Mary Douglas's "Where there is dirt there is system"). There are also such miscellaneous bon-bons as the unbotched version of what Neil Armstrong was meant to have said when he set foot on the Moon ("It's one small step for a man...") and Niels Bohr's "Predictions can be very difficult — especially about the future", which is usually attributed to the great American baseball player and malaprop-artist Yogi Berra. Inclusiveness is good, although sometimes the criteria for inclusion puzzled me. What is supposed to be relevant to science about Simone de Beauvoir's observation that "To be a woman, if not a defect, is at least a peculiarity"?

Dictionaries of quotations do not — and should not — have a point of view about their subject matter, so their higgledy-piggledy character displays the incoherence of equally compelling statements about practically any aspect of science. In contrast to Robert Merton's "Science is public, not

private knowledge", we have the assertion of his Harvard contemporary Percy Bridgman that "Science is essentially private". The mathematician David Hilbert insisted that neither science nor technology have "anything to do with the other", but, not many years later, D'Arcy Thompson observed that "When I was young, science walked hand-in-hand with art; now she walks arm-in-arm with trade".

Actually, these examples give a misleading impression of the dictionary. Many of the entries are neither memorable nor concise: philosopher Daniel Dennett is given 250 words to pronounce on human consciousness; Francis Crick has more than 270 to characterize the 'central dogma' of molecular biology; Pierre Duhem's discourse on the differences between British and Continental physics weighs in at almost 400 words; and the selection from Robert Hooke's "method of experimenting" is even longer than that. The all-encompassing *Oxford Dictionary of Quotations*, by comparison, has far fewer entries approaching that length.

But that's partly an artefact of the pool of material with which the editors of the *Oxford Dictionary of Scientific Quotations* had to work. The age of the scientist-aphorist probably ended with that of the scientist-generalist, the scientist-philosopher and the scientist-moralist, whose last great twentieth-century representatives included Albert Einstein, Robert Oppenheimer, J. B. S. Haldane, J. D. Bernal and Stephen Jay Gould, all of whom could turn a phrase. You don't just need literary skills to make a memorable statement: you also need a subject of some scope and cultural resonance, but the specialization of science and its ever-increasing divorce from the humanities make it harder for modern scientists to say resonant things in a resonant manner.

Nevertheless, the dictionary is a good source material, richly produced in the best Oxford University Press manner. It is one that many scientists will want to own, and will doubtless be ransacked for future Nobel prize acceptance speeches. Quibbles? Only a few: why no words by Dwight D. Eisenhower on the military-industrial complex, and no Alvin Weinberg on 'big science'? What about blunders? I noticed only one big

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Master of the *bon mot*: Oscar Wilde, without whom no dictionary of quotations could be complete, painted here by Henri de Toulouse-Lautrec.

one: the editors follow Lewis Wolpert in mistakenly attributing to Oppenheimer a statement about the social responsibility of the scientist that was in fact made by the man who destroyed Oppenheimer's reputation, Edward Teller. ■

Steven Shapin is Franklin L. Ford professor of the history of science, Harvard University, 1 Oxford Street, Cambridge, Massachusetts 02138, USA.

A natural history of India

Inventing Global Ecology: Tracking the Biodiversity Ideal in India, 1947–1997

by Michael L. Lewis

Ohio University Press: 2004. 352 pp.

\$55, £36.95 (hbk); \$26, £17.50 (pbk)

Deepak Apte

The title of this book couldn't be more apt. It is hard to be sure whether the word 'tracking' is a cleverly crafted metaphor or is just used to mean 'tracing the path of something'. Nevertheless, the word conjures up visions of an English gentleman (preferably with the name Major Bill Witherspoon) in a pith helmet, shielding himself from thorny acacia branches, extricating a 1948 Morris Minor from a mud pool, and trying to have a conversation in Swahili while pursuing a herd of *Loxodonta africana* (to 'shoot' them one way or the other) over the intractable terrain of sub-Saharan Africa. Or if you would prefer a less colonial scene, you might think of a young Indian postgrad trying to fulfil the specifications of her World Bank grant by pursuing the flighty (and nocturnal) Nilgiri marten as it scampers across the rainforest of the Kalakad Tiger Reserve. Or you might want to indulge in visions of technological empowerment and imagine instead a technocrat lounging at his computer station logging satellite data that beep in by the minute from a piece of silicon embedded within the dorsal skin of a brown palm civet.

Michael Lewis has done pretty much all of the above. He has faithfully traced and eloquently recorded the labyrinthine evolution of wildlife science (now called conservation biology). From its governing ideology and practice in India, he moves through the highways and byways (many of them literal, rather than figurative) of history, international politics (and politicking), Indian nationalism and nation-building. He even takes in the 'anthropology of science', the changing face of global scientific debate, and the evolution of field biology itself.

He has straddled the academic worlds of both biology (as an undergraduate) and history (graduate and doctoral). It therefore

seems natural that he should apply one to the other, and thus magnify insights into both. After all, they are intellectual disciplines located not separately in a vacuum, but in an interactive matrix of other fields of enquiry.

Inventing Global Ecology is based on Lewis's doctoral dissertation, so it reads as a fairly scholarly piece of writing. It smacks of painstaking research, which is its greatest claim to credibility. In fact, the information is often so detailed that the reader senses impatience within.

More importantly, the reader gets the distinct impression that Lewis has made a sincere attempt to embody 'deep democracy'. He is politically correct in the most profound and meaningful sense. He has the ability to transcend his inner barriers of rank, race, culture, affluent 'developed world'-ness and geopolitics. He writes with humility, grace, honesty and openness about the geopolitical realities of a people and its wildlife that are positioned often antagonistically and disadvantageously to his own. It is people such as Lewis, and the debate that books like this can initiate, that have the potential to make a real difference to difficult and chronic global problems.

The prose is smooth, unselfconscious and eloquent. For example, he uses elegant metaphors while writing about the Bombay Natural History Society (BNHS): "Perhaps more than a gateway, the BNHS has been a colossus astride Indian conservation-oriented ecology over the last fifty years." He also demonstrates his insight into the society and its activities. Of Salim Ali, its driving force for many years, Lewis writes: "His story is the story of the BNHS". He perceptively explains that ecology "was not introduced to India by

Westerners, but by the degreeless Ali, by way of his German training". It's certainly true that Ali brought the BNHS to the highest possible standards of ecological research in India. But it is because of Ali's pioneering work on Indian ornithology that even now the BNHS is known as an institute specializing only in birds.

Lewis also examines the various reasons for the shift in funding for the BNHS from Britain to the United States. Whether this shift was strategic or coincidental, it is worth adding that, in the past decade or so, the funding for BNHS projects has come mainly from Britain, completing the circle.

In the subsequent chapters, Lewis chronicles the evolution of the BNHS and various controversies during this process. Without being judgemental, he covers various facets of these debates in lucid prose. These chapters will be essential reading for anyone with an interest in the BNHS.

Chapter 2 focuses on what could be called the anthropology of science — the social, cultural and political constructs upon which science is based and conducted. There is a growing dialectic among social scientists around the related issues of what constitutes science, whether or not its nature and practice are independent of the context in which it exists, and on the appropriateness of transplanting scientific practices from one part of the world to another. Lewis uses this as a springboard to question whether the import of scientific techniques, legislative measures (such as prohibiting human habitation in wildlife sanctuaries) and methods of enquiry (for example, using expensive radio collars in developing countries) in wildlife science and management is meaningful or

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

On the move: Indians have made huge strides to improve their own conservation efforts.

R. SEITRE/STILL PICTURES

useful. This discussion is pertinent not only to wildlife science but to all spheres of enquiry.

Lewis does not mince his words. For instance, he observes that neither Raghavendra Gadagkar of the Centre for Ecological Sciences in Bangalore nor US ecologists such as E. O. Wilson “are willing to abandon theoretical ecology to take up taxonomy though. Gadagkar encourages students to do so, and Wilson ecologists of the developing world to do the same. Neither will do it himself.” And while talking about differences between India and the West, he writes that villages in India, “even those who lose crops and occasionally relatives, have existed in close contact with dangerous wildlife for centuries without driving those animals to extinction. Before the British

there was no evidence of Indian ‘predator elimination hunts’ in the style of the Texas rattlesnake drives or the turn-of-the-century wolf bounties in the United States, aimed at the eradication of every member of a given species.” There is something irresistibly refreshing about Lewis’s willingness to speak his mind.

In keeping with his seemingly earnest desire to faithfully reproduce all his findings, all of his conversations (or rather, the relevant sections of them) seem to be recorded verbatim. The book is thus peppered with quotes in ‘Hinglish’, and misplaced modifiers, fractured sentence construction and other endearing linguistic idiosyncrasies are rife.

Despite its scholarly character, this book is user-friendly for lay-readers, thanks in no

small measure to the author’s superb storytelling ability. The more informed reader might skip over a couple of pages here and there with a ‘been there, done that’ shrug, but I suspect that few ecologists are privy to the intricate history of their chosen field of work. The book thus serves to round out their education by providing a historical and philosophical perspective. ■

Deepak Apte is head of the Conservation Department, Bombay Natural History Society, Hornbill House, S. B. Singh Road, Mumbai 400 023, India.

Erratum

We apologize to Gordon M. Burghardt for misspelling his name in the recent review of his book *The Genesis of Animal Play* (*Nature* **434**, 273–274; 2005).

Science and culture

Science and superstition

Thomas Heatherwick’s sculpture for the Wellcome Trust’s new building in London.

Martin Kemp

Foyer sculptures — those big lumps of arty stuff commissioned to hang or stand in the entrance spaces of large modern buildings — don’t tend to occupy a high rank in the pantheon of contemporary art. There is often an inverse relationship between the size of the object and the attention it commands from those who come and go. Generally decided by committees, which are inherently uncreative, these sculptures typically occupy an unsatisfying middle ground: just modern enough to seem adventurous without imposing themselves demandingly on the viewer.

In this context, it is good to see that the massive piece commissioned from Thomas Heatherwick, which adorns a seven-story space in the Wellcome Trust’s new headquarters in London (designed by Hopkins Architects), is a technical, visual and conceptual tour de force.

The statistics are impressive enough in themselves: there are almost 150,000 glass spheres; 26,732 stainless steel wires, 0.5 mm in diameter, and the same number of springs; the sculpture is nearly 30 metres in height; and it has a total weight of 14 tonnes. Needless to say, such a work is not the product of an archetypal artist working in romantic isolation in a studio-garret. The teamwork required is akin to that in the workshop of a major Renaissance artist-engineer.

Even more impressive, for those who enter, leave and pass by the building, are the beguiling visual effects and complex associations embedded within the sculpture. Each of the glass beads is composed of two hemispheres surrounding a piece of dichroic film that generates a range of floating effects, from clear to turquoise, pink, green, yellow, violet and orange. The glass and film together ensure that the visual impact is continually transformed as the light changes

and the spectator moves through the space.

The overall form is the result of a self-organizing process. Heatherwick and his team experimented with a range of viscous molten substances, dropping them into water and watching them solidify in shapes that are unpredictable yet observe certain material parameters. Such configurations, suggestive of organic entities, were among those considered by D’Arcy Thompson in his great 1917 book, *On Growth and Form*. Metal proved most amenable, and, after some 400 tests, a small piece of solidified white metal was selected. This small model was then translated digitally by a screen matrix into an overall configuration of suspended beads.

The work that arose from this process does not have one single meaning. The parameters of interpretation are set in part by its presence in the headquarters of one of the world’s major medical research charities. Parallels of the structure with molecular models are almost inescapable.

The German title, *Bleigiessen*, has strong associations with health. ‘Lead-guessing’ is a New Year’s Eve ritual in central Europe. Molten lead is poured into water and the resulting shapes are divined as signalling a person’s fortune. Yet Heatherwick did not intend this as a meaning from the beginning. It was only late in the process of design that he learnt about ‘lead-pouring’ from his German grandmother.

The end result is that the inherent unpredictability of the process of cooling molten metal in water

belongs simultaneously to modern science and to folk intuitions about the vagaries of fortune that govern all our lives. Henry Wellcome, founder of the trust, would have been delighted by this unanticipated conjunction of modern scientific knowledge with the superstitions behind traditional customs.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK, and co-director of Wallace Kemp Artakt.



S. SPELLER

The Janus face of Mnemosyne

Memory: some systems in the brain may be better equipped to handle the future than the past.

Yadin Dudai and Mary Carruthers

"Time present and time past / Are both perhaps present in time future, / And time future contained in time past." Neuroscientists studying memory today might find T. S. Eliot's pairing in *Burnt Norton* of time past with time present easier to comprehend than his metaphysical linkage of times past and future. Modern research considers memory as a trace of past experience and, despite being equipped with the marvels of cellular biology and functional neuroimaging, most investigators are still looking for the brain equivalent of Plato's etched-wax tablet. But memory has not always been viewed in this way. It may prove useful to refresh our collective memory and reconsider a different view of memory — that which focuses on time future.

Shaped by the dominant reductionist paradigm, the term 'memory' in current neuroscience literature refers to processes ranging from modified reflexes in slugs to the recollection of Shakespeare's verses in humans. Furthermore, the terms 'neural plasticity' and 'memory' are often treated interchangeably, although the first reflects general hardware mechanisms that enable brains to adapt to change and the latter reflects specific use-dependent information subserved by that hardware. In this essay, 'memory' refers only to that faculty known in lay terms as 'recollection', and in slightly more professional vocabulary as 'episodic memory' or 'mental time travel'.

We all know from personal experience and introspection (research tools once cherished and now mostly disrespected) that mental excursions can be made not only to the past, but also to the future. It was even speculated (for example, by Endel Tulving) that the ability to contemplate future scenarios was a driving force in the evolution of episodic memory. This proposed selection-for-imagination might even be blamed for the inherent fallibility of our recollective faculty. Despite this, attempts to liken imagination to memory are still considered shaky by many; the latter is considered factual data (which it is not), the former the subject matter of poets and an unsubstantiated, inferior form of knowledge.

In ancient Greek mythology, Memory (Mnemosyne) was the mother of all the muses. In the mortal world, Aristotle, Galen, and their medieval Arab commentators, emphasized the role of memory in the ethical virtue of 'prudence', the ability to make wise

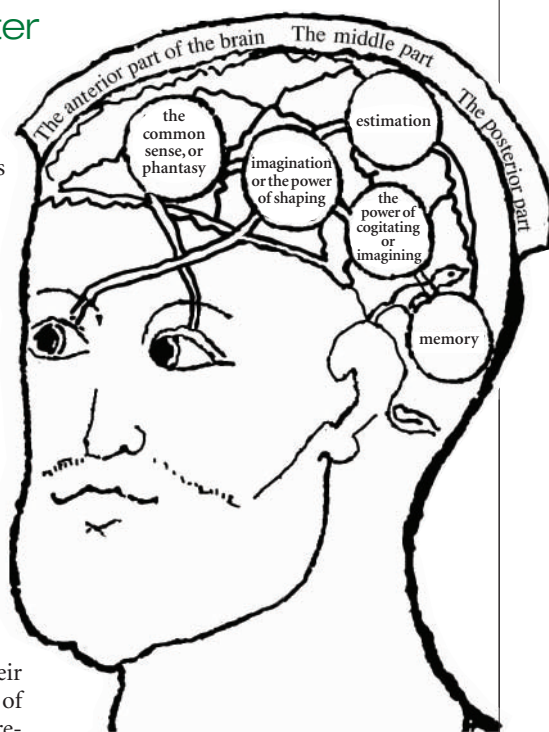
judgments and plan effectively. The word used by these scholars for concepts was 'phantasms' (in Greek, *phantasiai*; in Latin, *imagines*). Memory was also associated with prophetic writing, as in the Book of Ezekiel, whose prophecy consisted in recreating imaginatively the dimensions of the destroyed Temple in Jerusalem to envision and motivate the future — in this case, the Jews' return from captivity. Important in this planning effort is not accuracy of reproduction, but the act of imaginative recreation itself as a totally sensed and felt experience.

Reproduced to the right is a fourteenth-century analysis of the production of thought images, in which there is a central role for imagination. In this, memories are not direct signature-traces that past experiences leave in the brain. Their function is not to provide an accurate trace of things past, but to provide materials for creative thinking. In medieval analysis, data from all the senses came through various channels into the brain and were received in the 'phantasy' or 'common sense' area. The impressions were simultaneously collected and formed into images by the activity of 'making forms'. The experience also was 'estimated' emotionally as benign or hurtful, a gut reaction part of the mental imago. The resulting 'phantasm' was thus both a product of imagination and also had definite emotional colour; there was no such thing as a fully neutral or objective conception, and virtually all thoughts were formed as complete mental episodes.

Thought ('cogitation') was also an imaginative activity, and the activity of memory inventoried these products in a way that made them accessible to the recreative, investigative action of remembering. As there is a two-way relation between thinking and remembering, a kind of mental valve, the vermis, was thought to regulate the traffic. Many people lower their heads when thinking (as in Auguste Rodin's *The Thinker*) and raise them skyward when trying to recall a memory — this was considered to be engaging the action of the vermis.

Of course, it is not our aim to promote the erroneous brain biology of times past. But we do wish to revitalize the intuition and insight of scholars who didn't have powerful tools such as functional magnetic resonance imaging, and yet still wisely contemplated the global picture of the human mind. Their insights belong not only to the history of ideas, but are also relevant to neuroscience today.

Considering memory solely as a brain imprint of the past might limit the creativity



Medieval thoughts on the working of the brain.

of research programmes and bias the interpretation of their outcome. For example, experiences are correlated with changes at multiple levels of brain organization. Should such changes be considered solely as the impression of past experience, or also as the creation of the capacity to prepare for the future? Or another example: according to textbook accounts, memories consolidate into stable items shortly after their encoding. But do episodic memories ever really consolidate, given that our mind reshuffles them in playing and replaying past and future scenarios? Jean-Baptiste Molière would have been pleased to learn that some memory scientists may have spoken 'imagination-ish' all their life without even knowing it. Whether in the classroom or the lab, it is worth remembering that Mnemosyne has a Janus face, looking to both time past and time future. ■

Yadin Dudai is in the Department of Neurobiology at the Weizmann Institute of Science, Rehovot 76100, Israel, and a visiting Global Distinguished Professor of Neuroscience at New York University, New York 10003, USA. Mary Carruthers is at the Faculties of English, History and Religious Studies, New York University, New York 10003, USA.

FURTHER READING

Bolzoni, L. *The Gallery of Memory* (Univ. Toronto, 2001).
Carruthers, M. *The Book of Memory* (Cambridge Univ. Press, 1990).
Dudai, Y. *Memory from A to Z* (Oxford Univ. Press, 2002).
Tulving, E. in *The Missing Link in Cognition* (eds Terrace, H. S. & Metcalfe, J.) 3–56 (Oxford Univ. Press, 2005).

DNA search and rescue

Sheila S. David

How do DNA-repair enzymes find aberrant nucleotides among the myriad of normal ones? One enzyme has been caught in the act of checking for damage, providing clues to its quality-control process.

DNA-repair enzymes amaze us with their ability to search through vast tracts of DNA to find subtle anomalies in the structure. The human repair enzyme 8-oxoguanine glycosylase (hOGG1) is particularly impressive in this regard because it efficiently removes 8-oxoguanine (oxoG), a damaged guanine (G) base containing an extra oxygen atom, and ignores undamaged bases. Verdine and colleagues now report the structure of hOGG1 bound to undamaged DNA (Banerjee *et al.*¹, page 612 of this issue), revealing a unique strategy that allows faithful removal of damaged bases but not their normal counterparts.

The information content of the DNA double helix is preserved by a crew of DNA-repair enzymes that defend the genome from the harmful effects of DNA damage². A specialized pathway known as base-excision repair (BER) plays a primary role in rectifying base damage^{3,4}. The damaged base is removed by BER glycosylases, followed by excision of the remaining sugar fragment and installation of an undamaged nucleotide by a DNA polymerase. Notably, inherited defects in BER have recently been linked to colorectal cancer⁵.

A variety of agents cause oxidative damage to DNA, including oxygen radicals and ionizing radiation. Oxidation of G to form oxoG produces a subtle structural transformation that results in deleterious mutations because DNA polymerases misread oxoG as a thymine (T) base when the genome is being duplicated during cell division⁶. The human oxoG repair enzyme (hOGG1) catalyses excision of oxoG in the first step of BER. Structural studies of glycosylases involved in the repair process reveal common features of damaged-base recognition that include enzyme-initiated DNA

distortion and bending to flip the damaged base out from the DNA double helix for recognition within a base-specific cavity of the enzyme^{7–9}.

Verdine and colleagues previously determined the structure of an inactive hOGG1 variant bound to DNA containing an oxoG–cytosine (C) base pair¹⁰. Surprisingly, the only obvious mechanism by which hOGG1 discriminates between oxoG and G is through a single hydrogen bond to oxoG. It seemed unlikely that this single interaction would be sufficient to allow discrimination between the two bases, particularly given the million-fold excess of G over oxoG within the human genome. In addition, hOGG1 makes extensive contacts with the orphaned cytosine base, which ensures that oxoG is removed only when in the appropriate base-pairing context. Although extensive biophysical and structural studies intimate that there are general features of damaged bases that signal their presence to repair enzymes, the steps involved in finding damaged bases in a sea of normal ones are still unclear^{7–9}. Most mechanisms invoke the enzyme sliding or hopping along the DNA duplex until a damaged site is detected. A particularly intriguing question is whether normal bases are also extruded from the helix during the search process.

Banerjee *et al.*¹ reasoned that a structural snapshot of the enzyme encountering a G should provide insight into how the normal base is distinguished from its damaged oxoG counterpart. However, because the enzyme does not specifically recognize G, encountering and rejecting G is a transient process. Indeed, much like a train that stops only at certain locations, hOGG1 probably pauses only when close to oxoG. To capture the enzyme at an undamaged G–C base pair, the authors used an innovative ‘covalent

trapping’ method that they had previously devised to reveal the high-resolution structure of an HIV enzyme bound to DNA¹¹. Adding a cysteine thiol moiety in the enzyme and a single thiol-modified base in the DNA promoted the formation of a crosslink — a disulphide bridge — between the two when the enzyme reached the base. Banerjee *et al.* chose the appropriate site for the crosslink from the structure of inactive hOGG1 bound to an oxoG–C duplex¹⁰. Using the corresponding duplex containing G rather than oxoG and the trapping trick, they deduced the remarkable structure of hOGG1 bound to the undamaged DNA duplex.

Surprisingly, even though G is forcibly presented to the enzyme and is flipped out from the helix, it does not gain access to the pocket in the enzyme that binds oxoG, but to another site (Fig. 1). Calculations of differences in free energy indicate that both favourable and unfavourable interactions lead to preferential binding of oxoG over G in the oxoG-recognition pocket, and of G over oxoG in the alternative site. This structure captures an intermediate that forms in the process of finding oxoG, and illustrates that the damaged base must pass through a series of ‘gates’, or checkpoints, within the enzyme; only oxoG satisfies the requirements for admission to the damage-specific pocket, where it will be clipped from the DNA. Other bases (C, A and T) may be rejected outright without extrusion from the helix because hOGG1 scrutinizes both bases in each pair, and only bases opposite a C will be examined more closely.

Time-resolved fluorescence studies of the related DNA glycosylase that removes uracil (UDG) have provided evidence for the formation of a uracil intermediate partially removed from the DNA stack that forms

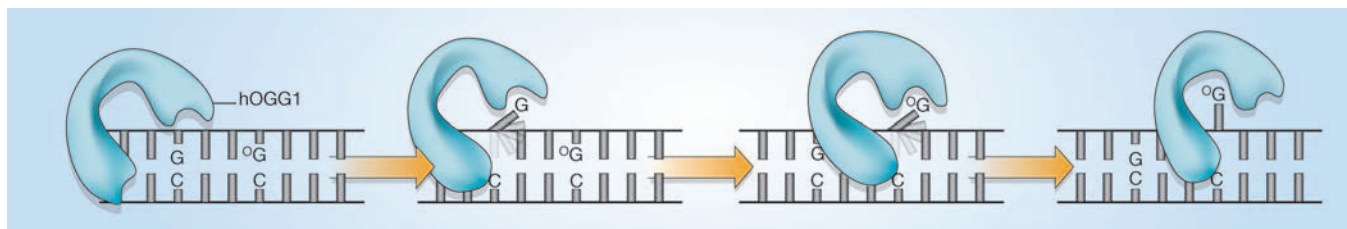


Figure 1 Damage detection — how hOGG1 searches for damaged guanine residues (oxoG, or °G), based on the findings of Banerjee and colleagues¹. The enzyme binds nonspecifically to the DNA and encounters a guanine–cytosine (G–C) base pair. Contacts with the C base results in extrusion of the G into the G-specific pocket.

The extruded G is denied further progression to the oxoG pocket, and goes back into the DNA double helix. When hOGG1 encounters an oxoG–C base pair, the oxoG is first extruded into the G-specific pocket, and then moves into an oxoG-specific pocket before being clipped out of the DNA.

before the uracil is fully extruded⁸. A uracil analogue that is unable to hydrogen-bond within the uracil-specific pocket may mimic this intermediate¹². Together, these studies^{1,8,12} suggest that a multi-step base-recognition process is a common strategy used by BER glycosylases to guarantee accurate base excision. Exploitation of Banerjee and colleagues' covalent trapping strategy may result in the detection of other intermediates in this process. In particular, clamping hOGG1 in the proximity of an A–T base pair may give insight into how it is dismissed by the enzyme. ■

Sheila S. David is in the Department of Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, Utah 84112, USA.

e-mail: david@chem.utah.edu

1. Banerjee, A., Yang, W., Karplus, M. & Verdine, G. L. *Nature* **434**, 612–618 (2005).
2. Schäfer, O. D. *Angew. Chem. Int. Edn Engl.* **42**, 2946–2974 (2003).
3. Barnes, D. E. & Lindahl, T. *Annu. Rev. Genet.* **38**, 445–476 (2004).
4. David, S. S. & Williams, S. D. *Chem. Rev.* **98**, 1221–1261 (1998).
5. Al-Tassan, N. et al. *Nature Genet.* **30**, 227–232 (2002).
6. Micheals, M. L. & Miller, J. H. *J. Bacteriol.* **174**, 6321–6325 (1992).
7. Fromme, J. C., Banerjee, A. & Verdine, G. L. *Curr. Opin. Struct. Biol.* **14**, 43–49 (2004).
8. Stivers, J. T. *Prog. Nucleic Acid Res. Mol. Biol.* **77**, 37–65 (2004).
9. Parikh, S. S., Putnam, C. & Tainer, J. A. *Mutat. Res.* **30**, 183–199 (2000).
10. Bruner, S. D., Norman, D. P. G. & Verdine, G. L. *Nature* **403**, 859–866 (2000).
11. Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. *Science* **282**, 1669–1675 (1998).
12. Jiang, Y. L. et al. *Biochemistry* **43**, 15429–15438 (2004).

Techniques

Spectroscopy at a stretch

Robin M. Hochstrasser

Two-dimensional spectroscopy can now be done using visible light. This allows the electronic couplings between energy levels to be measured directly and sheds new light on how molecules function in photosynthesis.

How light is converted into chemical energy in a photosynthetic system is largely determined by the vibrational and electronic dynamics of the complex biological macromolecules involved. An essential step in the elucidation of these mechanisms is the combined determination of structure and dynamics. The extension of nuclear magnetic resonance (NMR) spectroscopy into multiple dimensions¹ was introduced not long ago as one of the great advances in the study of the structure of proteins in solutions. Now, Fleming and co-workers², writing on page 625 of this issue, have transferred these principles of multidimensional spectroscopy to the electronic spectrum, where the wavelengths are about 10^{-7} of those in the NMR regime. This has allowed the first direct measurement of electronic couplings combined with the dynamics of excitations transferring between molecular energy levels in photosynthetic antennas.

The dimension of a spectroscopy refers to the number of independently variable time intervals between the field pulses that induce the signals. For example, with three femtosecond laser pulses there are three time intervals of interest — that between the first and second pulses (τ , the coherence time); that between the second and third (T , the waiting time); and the interval (t , the detection time) between the third pulse and the detected photon-echo field emitted by the sample in response to all three pulses. The three-dimensional grid of experimental data points obtained by varying these time intervals then defines a response in the time domain that can readily be converted to a

three-dimensional spectrum in the frequencies ω_r , ω_T and ω_t by Fourier transformation along the three time axes. The unique aspect of such spectra is that they exhibit cross peaks at spectral points ($\omega_r = \omega_A$, $\omega_t = \omega_B$) linking two transitions between molecular energy levels at the frequencies ω_A and ω_B . These cross peaks are present only because the structural components A and B of the system are able to sense one another's presence. In other words, A and B are close enough to be coupled and the pulse sequence must be inducing transitions between them. This information is not exposed directly by spectra in one dimension.

Multi-dimensional spectroscopy showing coupling between different structural units has also been demonstrated for electron paramagnetic resonance³ spectroscopy, for which the wavelengths used are much smaller than in NMR but still in the centimetre range, where the phase is relatively easy to control. More recently, experiments analogous to those using NMR have been performed in the infrared region of the spectrum⁴ at a wavelength of 6 μm , enabling the coupling between molecular vibrations to be exposed. Two-dimensional infrared spectroscopy has now become an active field of theoretical⁵ and experimental research, with applications in unravelling the dynamic structures of peptides and proteins^{6–10}, and liquids¹¹.

Fleming and colleagues² introduce two-dimensional spectra at a wavelength near the visible range (0.8 μm). The spectra are used to spread electronic transitions into two dimensions and expose the pattern of electronic couplings between the seven

bacteriochlorophyll groups whose absorption characteristics give these complex molecules their typical colour, and which correspond to the cofactors of a photosynthetic light-harvesting protein. Fleming and colleagues' spectra of ω_r versus ω_t over a range of waiting times change with T because energy is transferring between the cofactors as the system tends to equilibrium in its electronically excited state. The new method allows the energy transfer to be tracked on a cofactor-by-cofactor basis, providing a picture in space and time of the reorganization of the energy between the bacteriochlorophyll molecules in the light-harvesting complex of the bacterium *Chlorobium tepidum*.

The experiment clearly reveals which of the cofactors are coupled, and also yields the timescales of the energy flow between them. It is notable that the experimental two-dimensional electronic spectrum can be reproduced by a numerical analysis based on the electronic wavefunctions of the cofactors and a quantum theory of the interaction of cofactors with each other and with the protein environment. Theory and experiment combined have thus generated a molecular-scale map of the energy reorganization occurring within the light harvester, allowing us to visualize how it feeds energy to the reaction centre.

Because of its relatively broad bandwidth, this experimental method will not be restricted to the measurement of equilibrium dynamics. The pulse sequences that generate the photon echo could be applied to a system that has been kicked into a non-equilibrium state by another laser pulse. The new two-dimensional electronic and infrared methods are structural probes of the kinetics of chemical changes over a wide range of timescales dictated only by the ability to introduce spatial or electronic delays between short laser pulses; in comparison, the range over which kinetic data can be acquired with NMR is limited at present to tens of milliseconds¹². Furthermore, with the new techniques, the distance scale of the electronic couplings that can be measured can reach tens of nanometres, much more than is possible with NMR or two-dimensional infrared, both of which operate only at distances up to about 1 nm.

The approach detailed by Fleming and co-workers² thus opens a new era of electronic spectroscopy — one in which we can expect to visualize combined space–energy relationships between interacting electronic systems at the nanometre scale in a wide variety of applications. Ultimately these approaches will replace conventional ultraviolet, visible and infrared spectroscopy, because not only do they contain the structural information of the conventional spectra but they also tell whether or not the structural elements responsible for the absorption bands are near each other. The

next stage in the advancement of this technology will be to access even shorter wavelengths at which the residues and backbones of proteins and the bases of DNA have their electronic transitions.

Robin M. Hochstrasser is in the Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.
e-mail: hochstra@sas.upenn.edu

1. Ernst, R. R. *Principles of Nuclear Magnetic Resonances in One and Two Dimensions* (Oxford Univ. Press, 1987).
2. Brixner, T. *et al. Nature* **434**, 625–628 (2005).

3. Gorcester, J. & Freed, J. H. *J. Chem. Phys.* **88**, 4678–4693 (1988).
4. Asplund, M. C., Zanni, M. T. & Hochstrasser, R. M. *Proc. Natl Acad. Sci. USA* **97**, 8219–8224 (2000).
5. Mukamel, S. *Annu. Rev. Phys. Chem.* **51**, 691–729 (2000).
6. Hamm, P. & Hochstrasser, R. M. *Pract. Spectrosc.* **26**, 273–347 (2001).
7. Zanni, M. T. & Hochstrasser, R. M. *Curr. Opin. Struct. Biol.* **11**, 516–522 (2001).
8. Ge, N.-H. *et al. J. Phys. Chem. A* **106**, 962–972 (2002).
9. Chung Hoi, S., Khalil, M., Smith, A. W., Ganim, Z. & Tokmakoff, A. *Proc. Natl Acad. Sci. USA* **102**, 612–617 (2005).
10. Kim, Y. S. & Hochstrasser, R. M. *J. Phys. Chem. B* doi:10.1021/jp0449511 (2005).
11. Asbury, J. B. *et al. J. Chem. Phys.* **121**, 12431–12446 (2004).
12. Mok, K. H. *et al. J. Am. Chem. Soc.* **125**, 12484–12492 (2003).

Evolutionary biology

Why sex is good

Rolf F. Hoekstra

According to a proposal put forward many years ago, sexual reproduction makes natural selection more effective because it increases genetic variation. Experiments now verify that idea — at least in yeast.

Most animals and plants are sexual, and in organisms that normally multiply asexually, such as microbes and some groups of fungi, sexual processes are rarely completely absent. Such a widespread process must have an essential function. To their embarrassment, however, evolutionary biologists have had great difficulties in finding a simple and general explanation. As they describe elsewhere in this issue, Goddard and colleagues (page 636)¹ have used twenty-first-century techniques to provide confirmation of an idea, first mooted in the nineteenth century, as to why sex is good.

Sexual reproduction involves the marriage of genetic material from two parents to form progeny that transmit new

combinations of paternal and maternal genes to their offspring. It has spectacular consequences for the biology of organisms that extend beyond its evolutionary essence — the generation of new genetic combinations. In a world without sex there would be no males and females, no flowers, no insects specialized in pollinating them, no extravagant colour and form like the peacock's tail; and much animal behaviour aimed at finding and selecting mates would not exist.

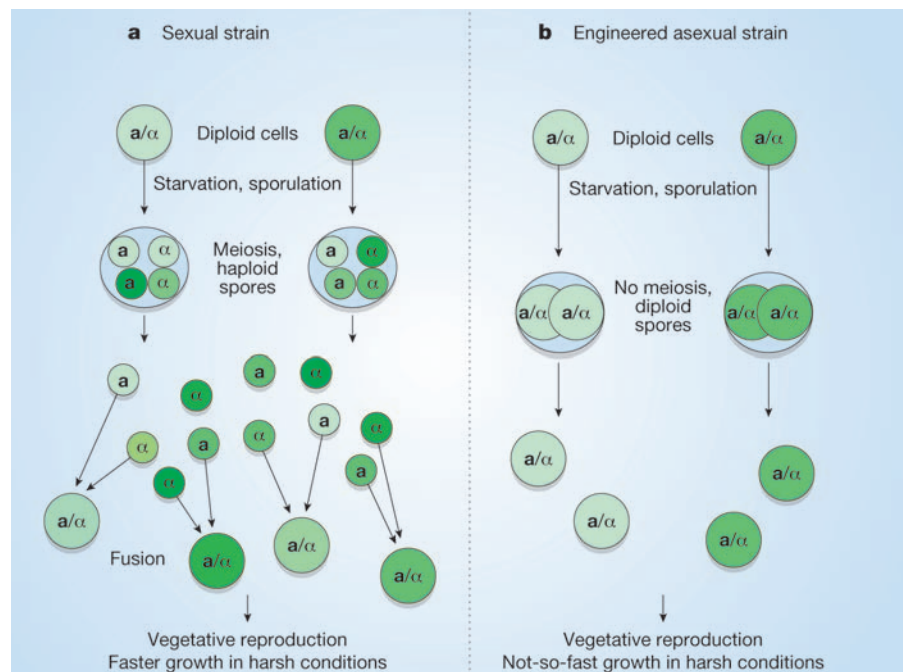
A large body of theory proposes a variety of hypothetical evolutionary advantages of sexuality and the genetic shuffling (recombination) that it involves^{2,3}. Discriminating between these theories empirically has proved very hard. But Goddard *et al.*¹ present the results of an elegant and rigorous experiment

with yeast, showing that a sexual population evolves faster than an asexual population when challenged by a novel environment.

When supplied with sufficient nutrients, yeast cells reproduce vegetatively (asexually), but if starved they undergo the process of meiosis, producing four spores that are functionally equivalent to the gametes involved in sexual reproduction. These spores then fuse in pairs to produce the next generation of vegetative cells (Fig. 1). Goddard *et al.* constructed an asexual yeast strain from a normal strain by deleting two genes (*SPO11* and *SPO13*) that are required for normal recombination and meiosis. This asexual strain differs from the sexual strain only in its ability to undergo meiotic recombination. When starved, the asexual strain forms two diploid spores, which are genetically identical to the parental cell, instead of the four recombinant haploid spores produced by the sexual strain.

The authors set out to test an idea, which traces back to August Weismann^{4,5} in the late nineteenth century, that sex increases genetic variation and thereby promotes evolutionary adaptation. Subjecting many generations of both the sexual strain and the engineered asexual strain to two novel environments, harsh and benign, they measured adaptation as increased 'fitness' (in this case, growth rate) relative to the non-evolved ancestral strain. In the benign environment, in which growth is limited by glucose concentration and where there is little selection, Goddard *et al.* observed no fitness increase in either the sexual or the asexual strains. However, in the harsh environment — having the same glucose concentration but at a higher temperature and with more demanding osmotic conditions — the sexual strain reached an increase in growth rate of 94% but the

Figure 1 Yeast reproduction, and Goddard and colleagues' experiments¹. Vegetative (diploid) cells proliferate asexually when food is available, but starvation induces sporulation; these diploid cells have both forms (a , α) of the 'mating type' gene. **a**, In the sexual strain sporulation involves meiosis, and so genetic recombination, and results in haploid spores that differ in genetic composition in mating type and elsewhere in the genome (indicated by the different colours). Pairs of spores with opposite mating type then fuse to produce a new generation of diploid vegetative cells. **b**, In the asexual strain, engineered to form unrecombined diploid spores but otherwise identical to the sexual strain, sporulation does not increase genetic variability. Here the diploid spores develop directly into the new vegetative cells. When both strains are grown in a harsh environment, the sexual strain shows faster evolutionary adaptation than the asexual strain, seen in terms of growth rate, owing to its greater genetic variability.



asexual strain reached only 80%. These results clearly support Weismann's hypothesis.

The history of ideas about the functional significance of sexual reproduction is remarkable. For a long time Weismann's theory was widely accepted. But in the 1970s several difficulties with his explanation were identified^{6,7}. First, genetic recombination will tend to break up favourable combinations of genes that have accumulated by selection, which may slow down adaptive evolution rather than speed it up. Second, in organisms with male–female differentiation, there is a so-called twofold cost of sex. If males do not contribute resources to offspring, a mutation to asexuality, causing females to produce only daughters, will — other things being equal — quickly increase in frequency. This is because sexual females produce only half the number of daughters compared to the asexual mutants, 'wasting' half their reproductive output on sons.

Decades of intense theoretical exploration ensued, aimed at discovering potential benefits of sex that are large enough to overcome these and other disadvantages. Many different hypotheses were advanced, most of them not mutually exclusive. The result was confusion and frustration rather than illumination. By 1993, hypotheses had become so numerous and diverse that a classification of them became necessary⁸. Since then, theoretical analysis has progressed further and has resulted in several plausible and general explanations, again not mutually exclusive, and roughly belonging to two categories: either sex increases the rate of adaptive evolution, or it is more efficient in eliminating deleterious mutations^{2,3}.

The solution to the evolutionary riddle of sex has to come from empirical tests of the various theories, and experimental studies

have been increasingly devoted to the problem⁹. However, clearcut comparisons between sexual and asexual reproduction are hampered by confounding factors. The paper by Goddard and colleagues¹ is exemplary in this respect, owing to the excellent possibilities offered by their yeast system.

Nonetheless, we are still far from a definitive answer to the question of why sexual reproduction is so common. For one thing, yeast has no male–female distinction, so the twofold cost of sex does not apply, which prevents a straightforward generalization of the conclusions to most animals and plants. Moreover, the experiment provides no information as to why the sexual population is more efficient in adapting to the harsh conditions than the asexual population. Is it because sex has brought together different beneficial mutations that have arisen in separate lineages, or because sex has liberated beneficial mutations from unfavourable genetic interactions with other genes? Further genetic characterization of the mutations responsible for the adaptation to the harsh environment would be a first step to distinguishing between these possibilities. ■

Rolf F. Hoekstra is in the Laboratory of Genetics, Wageningen University, Arboretumlaan 4, 6703 BD Wageningen, The Netherlands.
e-mail: rolf.hoekstra@wur.nl

1. Goddard, M. R., Godfray, H. C. J. & Burt, A. *Nature* **434**, 636–640 (2005).
2. Otto, S. P. & Lenormand, T. *Nature Rev. Genet.* **3**, 252–261 (2002).
3. Barton, N. H. & Charlesworth, B. *Science* **281**, 1986–1990 (1998).
4. Weismann, A. *Die Bedeutung der sexuellen Fortpflanzung für die Selektionstheorie* (Fischer, Jena, 1886).
5. Weismann, A. *The Evolution Theory* (Arnold, London, 1904).
6. Williams, G. C. *Sex and Evolution* (Princeton Univ. Press, 1975).
7. Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, 1978).
8. Kondrashov, A. S. *J. Hered.* **84**, 372–387 (1993).
9. Rice, W. R. *Nature Rev. Genet.* **3**, 241–251 (2002).

Aceh–Andaman earthquake

What happened and what's next?

Kerry Sieh

The huge earthquake of 26 December 2004 and ensuing tsunami were caused by a submarine rupture running from offshore Aceh, Indonesia, to the Andaman Islands. A clearer picture of events is starting to emerge.

As the human drama of the Aceh–Andaman earthquake and tsunami unfolded in the last days of 2004, laymen and scientists began scrambling to understand what had caused these gigantic disturbances of Earth's crust and seas. One of the earliest clues was the delineation, within just hours of the mainshock, of a band of large aftershocks arcing 1,300 km from northern Sumatra almost as far as Myanmar (Burma)¹. This seemed to signal that about 25% of the Sunda megathrust, the great tectonic boundary along which the

Australian and Indian plates begin their descent beneath Southeast Asia, had ruptured. In less than a day, however, analyses of seismic 'body' waves² were indicating that the length of the rupture was only about 400 km.

This early controversy about the length of the megathrust rupture created a gnawing ambiguity about future dangers to populations around the Bay of Bengal. If only 400 km of the great fault had ruptured, large unfailed sections might be poised to deliver another tsunami. If, on the other hand, most of the submarine fault had broken,



100 YEARS AGO

Philosophy as Scientia Scientiarum and a History of Classifications of the Sciences. The relation of science to philosophy is, in theory, filial. It is, perhaps, no contradiction of the filial relationship that in practice it has an unfortunate tendency to run to mutual recrimination. The man of science too often ignores the philosopher, or despises him as an obscurantist who habitually confounds abstraction with generalisation. To the metaphysical philosopher, on the other hand, the typical specialist in science is a variety of day-labourer, dulled by the drudgery of occupational routine... Prof. Flint's new book should serve as a mediating influence between philosophical and scientific interests. It brings together into one convenient source the leading attempts made, from Plato to Karl Pearson, towards a classification of the sciences... How invaluable a service Prof. Flint has thus rendered to future investigators, can be appreciated only by those who have tediously toiled at the scattered literature of this subject. Its bibliography appears hitherto to have been left unorganised — having escaped even the ubiquitous zeal of German scholarship.

From *Nature* 30 March 1905.

50 YEARS AGO

The Piltdown saga has now been concluded. The British Museum (Natural History) has brought together under the title "Further Contributions to the Solution of the Piltdown Problem"... the views of a large number of specialists who have described their work on the specimens — work which has resulted in demonstrating without any doubt that the whole thing was a hoax... Dr J. S. Weiner's book, "The Piltdown Forgery", deals rather with the many personalities in the case... The story unfolds with the discovery of Piltdown I, and we find not a few of the local amateurs a little suspicious that all is not as presented. We now have evidence for staining of specimens; we have the 'happy' find of Piltdown II which was to clinch the matter... It all reads like a novel, and I have no intention of spoiling it for others by saying more. It is a book to be read with interest and profit. Such a hoax nowadays is impossible; but it was one actually perpetrated and it was a great success.

From *Nature* 2 April 1955.



Figure 1 Evidence of uplift — here about 1.5 m. This aerial photo shows new land that emerged on the southwest coast of Simeulue island, above the southern edge of the megathrust rupture that caused the Aceh–Andaman earthquake. The wide strip of land consists of a former fringing coral reef; what was previously a beach (far left) has been left high and dry. Inset, a field of fire-coral heads on a neighbouring coral reef that rose about a metre during the earthquake.

then the chances of such a disaster were much smaller.

In this issue, Ni, Kanamori and Helmburger (page 582)³ explain why early analyses grossly underestimated the rupture length, and they present an analysis of high-frequency (2–4 Hz) seismic signals that clearly shows northward propagation of the rupture for a distance of about 1,200 km. Also in this issue, Stein and Okal (page 581)⁴ argue that early estimates of the magnitude^{1,2} were far too low. Using extremely long-period seismic ‘normal mode’ waves, they calculate that the earthquake’s magnitude was 9.3, about three times larger than initial estimates of 9.0 (given the logarithmic nature of earthquake-magnitude scales). This much larger size is consistent with slip averaging about 13 m along a 1,200-km rupture, assuming that much of the slippage occurred too slowly to be seen in shorter-wavelength seismograms. Thus, they claim the long-versus-short rupture controversy is solved and that there is no need to worry about another giant earthquake and tsunami originating along this long section of the fault.

These two reports^{3,4} are among the first published analyses of what is destined to be one of the most important earthquakes of the century. Over the next year or two, figuring out what happened will be a showcase both of what modern observations and analysis can do and of the multidisciplinary nature of modern earthquake science⁵. In the months ahead, much more will be learned about this giant event. Satellite imagery and field measurements of dramatically uplifted and submerged coastlines (Fig. 1)^{6,7} and the movement of Global Positioning System geodetic stations⁸, as well as tsunami records, will all add constraints on the areal extent of the rupture and the magnitude and sequencing

of slip: these, in turn, will be essential to understanding the tsunami.

If all of the megathrust between northernmost Sumatra and Myanmar has produced its once-a-millennium giant earthquake, why should we have any immediate concern about another giant quake or tsunami in the Bay of Bengal? McCloskey *et al.*⁹ offered one answer by estimating the stresses imposed by the giant 2004 rupture on the two big faults farther south. It seems that the section of the Sunda megathrust immediately to the south, off the coast of northern Sumatra, is now closer to failure. Likewise for the nearest portion of the great San Andreas-like Sumatran fault, which runs through Banda Aceh and down the backbone of the Sumatran mainland.

The critical question is how close to failure the 2004 rupture has moved these two big faults. This will be moot until more is known about the history of their past ruptures. It will be necessary to learn how the Sumatran parts of the megathrust are segmented structurally, and how they have behaved in the past. Immediately south of the 2004 rupture, for example, it appears from the historical record that there were very large earthquakes in 1861 and 1907¹⁰. Where on the megathrust were these ruptures, and how often and how regularly do they recur? Palaeoseismic data are available only for a 700-km-long section farther away, from about 1° to 5° south of the Equator. Giant earthquakes and tsunamis occur there about every 200–230 years, sometimes as a single giant earthquake, sometimes as two in relatively quick succession, as happened in 1797 and 1833^{11,12}.

Big faults on the northern flank of the 2004 rupture also pose a hazard; the northern extension of the 2004 rupture continues for another 1,000 km, up the west coast of

Myanmar, well past Bangladesh to the eastern end of the Himalayas. Too little is known of its long-term history to provide a meaningful assessment of its future behaviour. Moreover, long sections of the enormous thrust fault along which India is diving down beneath the Himalayas have not failed for centuries and are only one to three fault-lengths away from the 2004 rupture.

It is sobering to realize that big earthquakes sometimes occur in clusters (for example, seven of the ten giant earthquakes of the twentieth century occurred between 1950 and 1965, and five of these occurred around the northern Pacific margin)¹³. Because many of the giant faults in the Aceh–Andaman neighbourhood have been dormant for a very long time, it is quite plausible that the recent giant earthquake and tsunami may not be the only disastrous twenty-first-century manifestation of the Indian plate’s unsteady tectonic journey northward. ■

Kerry Sieh is at the Tectonic Observatory, California Institute of Technology, Pasadena, California 91125, USA.

e-mail: sieh@gps.caltech.edu

1. http://neic.usgs.gov/neis/poster/2004/20041226_image.html
2. www.gps.caltech.edu/~jichen/Earthquake/2004/aceh/aceh.html; www.eri.u-tokyo.ac.jp/sanchu/Seismo_Note (in Japanese).
3. Ni, S., Kanamori, H. & Helmburger, D. *Nature* **434**, 582 (2005).
4. Stein, S. & Okal, E. A. *Nature* **434**, 581–582 (2005).
5. Committee on the Science of Earthquakes *Living on an Active Earth: Perspectives on Earthquake Science* (Nat'l Academies Press, Washington DC, 2002).
6. www.tectonics.caltech.edu/sumatra/main/data.html
7. Bilham, R. Preprint at <http://cires.colorado.edu/~bilham>
8. www.seires.net/content/view/full/122/52
9. McCloskey, J., Nalbant, S. S. & Steacy, S. *Nature* **434**, 291 (2005).
10. Newcomb, K. R. & McCann, W. R. *J. Geophys. Res.* **92**, 421–439 (1987).
11. Sieh, K., Natawidjaja, D., Chlieh, M., Galetzka, J. & Avouac, J.-P. *Trans. Am. Geophys. Un. Fall Meet. Suppl. Abstr. T12B-04* (2004).
12. Zachariasen, J., Sieh, K., Taylor, F. W., Edwards, R. L. & Hantoro, W. S. *J. Geophys. Res.* **104**, 895–919 (1999).
13. Kanamori, H. *Nature* **271**, 411 (1978).

Cell biology

The more MAD, the merrier

Robert S. Hagan and Peter K. Sorger

Cells must pass the correct number of chromosomes to their progeny through the complex ballet of cell division. An unusual conformation-sensitive switch seems to maintain accurate chromosome segregation.

The maintenance of normal chromosome number during cell division requires the precise separation of duplicated sister chromosomes into two equal sets. Andrea Mussachio and colleagues¹, writing in *Current Biology*, shed light on how a protein called Mad2 ensures the fidelity of this process. The authors propose that two different structural forms of Mad2 collaborate to change one of the partners from an inactive to an active state. Such 'self-templated' changes in protein conformation are a characteristic of prions, but represent a new mechanism for cell signalling.

To achieve the correct separation, each chromosome is attached through the multi-protein 'kinetochore' complex to an intracellular scaffold called the spindle. In the 'anaphase' stage of cell division, this structure pulls the sister chromosomes apart to form two groups at opposite ends of the cell, each with one sister from each chromosome pair. The two groups will be surrounded by membranes to become the nuclei of the

daughter cells. The onset of anaphase is blocked until all chromosomes are bound correctly to the spindle. Defects in the proteins that make up this blockade (or 'spindle checkpoint') lead to abnormal chromosome numbers in the daughter cells, with potentially disastrous consequences — many tumour cells, for example, have abnormal chromosome complements. The spindle checkpoint activates a cascade of signals that converge on Cdc20, a protein that regulates the anaphase-promoting complex (APC)². If even a single kinetochore remains unattached to the spindle, a diffusible signal is generated that is sufficient to restrain Cdc20 and block cell division at the start of anaphase for many hours³. But the nature of the signal and the mechanism by which it is propagated remain elusive.

Two highly conserved Cdc20-binding proteins, Mad2 and BubR1, are essential to inhibit APC (and therefore anaphase). These attach to kinetochores early in cell division, and dissociate as the kinetochore attaches to the spindle, but they persist on the

kinetochores of misoriented or unattached chromosomes (reviewed in ref. 2). Cdc20, Mad2 and BubR1 cycle rapidly on and off kinetochores, suggesting that they might be part of the diffusible checkpoint signal^{3,4}.

Previous studies have focused on the ability of Mad2 to bind to Cdc20 and the checkpoint protein Mad1. The Mad1–Mad2 interaction localizes Mad2 to kinetochores and seems to occur early in the checkpoint pathway, whereas the binding of Mad2 to Cdc20 blocks APC activation and happens later in the checkpoint cascade (for review see ref. 2). Upon associating with Mad1 or Cdc20, the carboxy-terminal tail of Mad2 rearranges from a disordered 'open' conformation (OMad2) to a closed conformation (CMad2), thereby encircling Mad1 or Cdc20 peptides like a safety belt^{5,6}. Mad1 and Cdc20 compete for the same binding site on Mad2 but, paradoxically, Mad1 is required for the formation of Cdc20–Mad2 complexes *in vivo*. The high stability of Mad1–Mad2 complexes and the slow spontaneous inter-conversion of OMad2 and CMad2 seem at odds with the rapid kinetics of checkpoint signalling.

De Antoni *et al.*¹ use Mad2 mutants locked into open and closed conformations to show that two Mad2 molecules participate in this complicated dance. Mad2 variants lacking ten amino acids at the carboxy terminus cannot close the safety belt, are stuck in the open state and do not bind to Mad1 or Cdc20. However, the authors show that 'always-open' Mad2 can bind stably to Mad1–CMad2 complexes. Strikingly, OMad2 and CMad2 can partner each other, but two molecules of the same conformation cannot associate. In cells, the always-open OMad2 localizes to kinetochores when normal Mad2 is present, but it cannot engage the checkpoint, presumably because it cannot bind directly to Cdc20. Importantly, Mad2 mutants that cannot form partnerships with other Mad2 molecules do not bind to Mad1–CMad2, localize to kinetochores or support checkpoint function.

From these data, the authors propose a model in which 'core' complexes of Mad1–CMad2 increase the concentration and stability of OMad2 at kinetochores (Fig. 1). When the checkpoint is activated, the OMad2 in Mad1–CMad2–OMad2 complexes can be transferred to Cdc20, closing up to form Cdc20–CMad2. Such complexes prevent APC activation and thus arrest cell division. This is a striking shift from previous models in which just a single CMad2 molecule bound to Mad1 before unfolding its safety belt and transferring to Cdc20. This new model fits neatly with data showing that two distinct pools of Mad2 associate with kinetochores: one that remains stably bound throughout cell division, and one that continuously binds and dissociates with a half-life of about 20 seconds^{3,7}. As predicted by

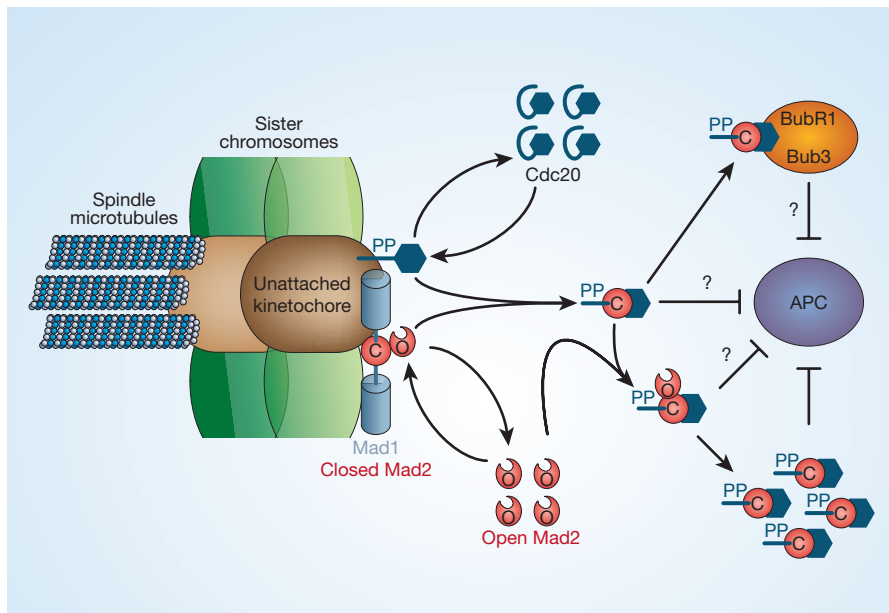


Figure 1 An asymmetric Mad2 partnership at the centre of the spindle checkpoint machinery. Mad2 encloses Mad1 to form a stable complex. Open-conformation Mad2 (OMad2) molecules bind to and are released from the Mad1–Mad2 complex in a cyclic fashion. At kinetochores (brown) that lack microtubule attachment, the OMad2 molecule encloses and binds activated Cdc20 (black), a step that may be promoted by the phosphorylation (PP) of Cdc20. The resulting closed Mad2 (CMad2)–Cdc20 complex inhibits anaphase-promoting complex (APC) and may bind BubR1 and Bub3. This complex may also act as a template for the formation of more CMad2–Cdc20 complexes from the cytosolic pool of OMad2, amplifying the checkpoint signal. Multiple protein complexes may be responsible for APC inhibition; their exact composition remains unknown.

the model, Cdc20 also cycles on and off kinetochores with similar kinetics, whereas Mad1 does not.

Despite the elegance of De Antoni and colleagues' model, many questions remain unanswered. It is not clear how kinetochore-spindle attachment regulates Mad2-mediated checkpoint signalling. Also unknown is the molecular nature of the diffusible signal; De Antoni *et al.* suggest that Cdc20–Cmad2 may constitute this signal and may act as an amplifier by binding to and activating other OMad2 molecules. Finally, the notion that Mad2 can inhibit Cdc20 only after being activated by Mad1 does not explain how cytosolic Mad2 regulates Cdc20 early in the cell cycle in a Mad1-independent manner⁸.

One promising approach to tackling these issues is to focus on Cdc20. Cdc20 is found in cells in association with Mad2, and two other proteins, BubR1 and Bub3 (ref. 9), and is phosphorylated at multiple sites. Intriguingly, Mad2 binds tightly *in vitro* to Cdc20 in the absence of any other proteins (most notably Mad1). *In vivo*, however, Mad2 requires kinetochores that are not attached to the spindle to bind to Cdc20, implying that access to the Mad2-binding site on Cdc20 is regulated. One possible reg-

ulator is BubR1, which is required for both the kinetochore-dependent and kinetochore-independent functions of Mad2. An attractive possibility is that all of these inputs regulate a transition in Cdc20 between conformations that are susceptible to Mad2-dependent inhibition and those that are resistant. Although much remains to be learned about Cdc20 and the spindle checkpoint in general, the work of De Antoni *et al.* highlights the central role of Mad2 in guiding a cell through cell division with all of its chromosomes in order.

Robert S. Hagan is in the Department of Biology, and Peter K. Sorger is in the Department of Biology and Biological Engineering Division, Massachusetts Institute of Technology, 77 Massachusetts Avenue, 68-371, Cambridge, Massachusetts 02139, USA. e-mail: psorger@mit.edu

1. De Antoni, A. *et al.* *Curr. Biol.* **15**, 214–225 (2005).
2. Rieder, C. L., Cole, R. W., Khodjakov, A. & Sluder, G. *J. Cell Biol.* **130**, 941–948 (1995).
3. Howell, B. J., Hoffman, D. B., Fang, G., Murray, A. W. & Salmon, E. D. *J. Cell Biol.* **150**, 1233–1250 (2000).
4. Musacchio, A. & Hardwick, K. G. *Nature Rev. Mol. Cell Biol.* **3**, 731–741 (2002).
5. Luo, X., Tang, Z., Rizo, J. & Yu, H. *Mol. Cell* **9**, 59–71 (2002).
6. Sironi, L. *et al.* *EMBO J.* **21**, 2496–2506 (2002).
7. Shah, J. V. *et al.* *Curr. Biol.* **14**, 942–952 (2004).
8. Meraldi, P., Draviam, V. M. & Sorger, P. K. *Dev. Cell* **7**, 46–60 (2004).
9. Sudakin, V., Chan, G. K. & Yen, T. J. *J. Cell Biol.* **154**, 925–936 (2001).

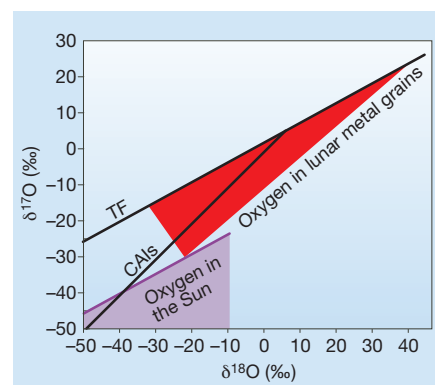


Figure 1 Oxygen in the Solar System. The quantities $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$ express deviations in the ratios $^{17}\text{O}/^{16}\text{O}$ and $^{18}\text{O}/^{16}\text{O}$ relative to the same ratios in 'Standard Mean Ocean Water' (SMOW) on Earth, where $\delta^{17}\text{O} = [(^{17}\text{O}/^{16}\text{O})_{\text{sample}} / (^{17}\text{O}/^{16}\text{O})_{\text{SMOW}} - 1] \times 1,000$ and $\delta^{18}\text{O} = [(^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (^{18}\text{O}/^{16}\text{O})_{\text{SMOW}} - 1] \times 1,000$. All terrestrial rocks and water lie along the line labelled TF. Refractory inclusions in meteorites lie along the line labelled CAIs. Most meteorites lie slightly below or above the TF line. Because oxygen isotopic compositions of lunar metal grains lie in the red field, and all lunar rocks lie along the TF line, the oxygen isotopic composition of the Sun is inferred to be somewhere in the light purple field, on or below the dark purple line.

Cosmochemistry

A breath of solar air

Andrew M. Davis

The abundance of different isotopes of oxygen provides clues to the origin of matter in the Solar System. Hence the importance of studies of oxygen atoms from the Sun trapped in metal grains on the Moon.

The Sun contains most of the matter in the Solar System, but surprisingly little is known about its isotopic composition. Spectroscopic methods can be applied only to a few abundant elements and are imprecise compared with laboratory measurements on Earth. The only practical way to sample the Sun is to measure the isotopic composition of matter emitted from it.

Such emissions can exist in one of two forms: 'solar wind' consists of relatively low-energy charged particles from the Sun's corona — mainly protons and helium ions, but also lesser quantities of every element — accelerated to speeds of 300 to 800 km s⁻¹; and the less abundant 'solar energetic particles' (SEPs), also comprising ions of all the elements, but with energies five to a hundred times greater than those of the solar-wind ions. Solar wind and SEPs can be implanted in materials exposed to the Sun in space or on the surface of planetary bodies without atmospheres, such as the Moon. In the latest in a series of studies using minerals in

lunar soils to infer isotopic compositions of elements in the Sun, Hashizume and Chaussidon (page 619 of this issue)¹ report that the Sun is enriched in the oxygen isotope ^{16}O compared with most objects in the Solar System.

But why is the abundance of oxygen isotopes so important? Oxygen has three isotopes, ^{16}O , ^{17}O and ^{18}O , that exist in the proportions 2,700:1:5. For a long time, only ^{16}O and ^{18}O were measured for geo- and cosmochemical purposes. This was because, in most physical and chemical processes, the fractionation of oxygen isotopes from one another is a mass-dependent process, with the $^{18}\text{O}/^{16}\text{O}$ ratio varying twice as much as the $^{17}\text{O}/^{16}\text{O}$ ratio. Thus, it was thought that the $^{17}\text{O}/^{16}\text{O}$ ratio in a sample could be inferred from the $^{18}\text{O}/^{16}\text{O}$ ratio. The situation changed in 1973, with the discovery² that calcium-aluminium-rich refractory inclusions (CAIs) in primitive chondrite meteorites are enriched in ^{16}O alone (Fig. 1). Since that discovery, oxygen isotopes have proven to be a powerful tool for studying the relations

between Solar System materials, such as planets, meteorites, chondrules and CAIs.

We now know the oxygen isotopic compositions of Earth, the Moon, Mars and the parent bodies of various kinds of meteorites. What we do not know is the composition in the Sun, where more than 99% of oxygen in the Solar System resides. The oxygen isotopic anomalies in Solar System materials were thought to be due to varying mixtures of presolar matter that gained their oxygen isotopic heterogeneity by nucleosynthesis in different kinds of stars. (Nucleosynthesis is the process by which elements are formed through nuclear reactions.) However, laboratory experiments suggested that the anomalies might be photochemical in origin^{3,4}.

In an influential paper, Clayton⁵ suggested that the anomalies do indeed arise through a photochemical effect known as self-shielding. In this model, ultraviolet radiation from the early Sun dissociates carbon monoxide — the most abundant oxygen-containing molecule in solar gas — in the inner part of the solar nebula, close to the Sun. The radiation of a wavelength appropriate to dissociate C^{16}O has a much shorter path length than that needed to dissociate C^{17}O and C^{18}O , because C^{16}O is so much more abundant. In the self-shielded portion of the nebula, only C^{17}O and C^{18}O would be dissociated. The resulting reactive ^{17}O - and ^{18}O -enriched oxygen atoms would then form silicates — a major constituent of most rocks — in an unspecified way.



Figure 2 Lunar sunrise. The Moon's lack of an atmosphere means not only that the Sun shines more brightly there than on Earth, but also that its surface is exposed to solar wind and solar energetic particles. This photograph was taken by the Apollo 12 mission of 14–24 November 1969.

The quantity $\Delta^{17}\text{O}$ (which is defined as $\delta^{17}\text{O} - 0.52 \delta^{18}\text{O}$) is used as a measure of the deviation of oxygen isotopic composition relative to the terrestrial fractionation line TF (Fig. 1), which describes the composition of terrestrial rocks and water. The most extremely ^{16}O -enriched composition, commonly found in CAIs, has $\delta^{18}\text{O}$ and $\delta^{17}\text{O}$ values of around -40% , leading to a value for $\Delta^{17}\text{O}$ of -20% . Clayton predicted that the Sun has this extreme composition, and that the isotopic make-up of all the meteorites and planets sampled has been modified by self-shielding close to the Sun; newly formed silicate dust would have been thrown out from this region by stellar winds.

Two alternative models, also involving self-shielding, have been proposed. One model⁶ is that oxygen isotopic heterogeneities observed in molecular clouds (and generated by self-shielding) were inherited and partially preserved during the formation of the Solar System. The other⁷ is that self-shielding occurred in surface regions of the accreting solar nebula at distances of several astronomical units (1 AU is the distance from the Earth to the Sun).

Hashizume and Chaussidon's findings¹ represent only one of several attempts to collect and analyse material of solar origin. Most notably, the Genesis mission, which returned to Earth rather abruptly in September 2004, exposed collectors made of high-purity materials, such as silicon, silicon carbide and diamond, to the solar wind for 30 months. Unfortunately, the collector materials are contaminated with soil, heat-shield material, pieces of other collectors and paint chips because of the failure of the parachute to deploy on re-entry and the resulting crash of the spacecraft into the Utah desert. The Genesis team is confident that these contaminants can be removed,

and initial measurements of solar composition are expected within a year.

Hashizume and Chaussidon also had to overcome a problem of contamination, but of a different kind. They investigated lunar soils containing iron and iron–nickel grains that had been exposed to the solar wind and SEPs for millions of years (Fig. 2). The relatively low-energy solar-wind particles are implanted in the outer 50–100 nanometres

of the lunar grains. But oxide coatings and submicroscopic silicate mineral grains on the surface of the metal grains prevented Hashizume and Chaussidon from detecting the solar-wind signal. However, in some grains the SEPs are implanted to depths of several hundred nanometres. In these cases, the surface oxide layer could be sputtered away and, in a few grains, isotopically anomalous oxygen was found with $\Delta^{17}\text{O}$ up to -20% .

Remarkably, this number is identical to that predicted by self-shielding models^{5–7} for the composition of the Sun; it will be very interesting to see if the Genesis collector materials give the same value. For now, however, it seems that it is not the Sun that has an abnormal isotopic composition: it is all the bodies we know in the inner Solar System — Earth, the Moon, Mars and the asteroids that are parent bodies to meteorites — that are enriched in ^{17}O and ^{18}O , by some 5%. ■

Andrew M. Davis is in the Chicago Center for Cosmochemistry, the Enrico Fermi Institute, and the Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois 60637, USA. e-mail: a-davis@uchicago.edu

1. Hashizume, K. & Chaussidon, M. *Nature* **434**, 619–622 (2005).
2. Clayton, R. N., Grossman, L. & Mayeda, T. K. *Science* **182**, 485–488 (1973).
3. Thieme, M. H. & Heidenreich, J. E. III *Science* **219**, 1073 (1983).
4. Thieme, M. H. *Science* **283**, 341 (1999).
5. Clayton, R. N. *Nature* **415**, 460–461 (2002).
6. Yurimoto, H. & Kuramoto, K. *Science* **305**, 1763–1766 (2004).
7. Lyons, J. R. & Young, E. D. *Lunar Planet. Sci.* XXXV, abstr. 1970 (2004).

Biochemistry

A pore way to die

Andrew Halestrap

Like Jekyll to Hyde, mitochondria can change from acting as the cell's powerhouses to become merciless killers. Mice lacking this mechanism develop normally, and their hearts are resistant to pathological damage.

As part of normal development and tissue turnover, cells die in a controlled, energy-dependent manner known as apoptosis. However, in response to noxious insults such as lack of oxygen, cells die by an uncontrolled, energy-independent process called necrosis. Necrosis culminates in rupture of the damaged cells, loss of their contents and a consequent inflammatory response that exacerbates injury¹. Starting on page 652 of this issue, Nakagawa *et al.*² and Baines *et al.*³ describe mice that lack a protein — cyclophilin D — thought to be critical in cell death, and provide insight into the mechanisms of both necrosis and apoptosis.

Necrosis occurs, among other instances, following a heart attack or stroke, when a clot disrupts the blood supply to an area of the heart or brain, and cells begin to die by necrosis. Restoring blood flow rapidly using clot-busting drugs or angioplasty allows the

tissue to recover, but after long periods without blood flow (ischaemia), reperfusion actually causes further necrosis and irreparable damage. A major cause of such reperfusion injury lies with the mitochondria, intracellular organelles that produce ATP (the 'energy currency' of cells). Necrosis involves the opening of a pore in the inner mitochondrial membrane, known as the mitochondrial permeability transition pore (MPTP)⁴. This process is triggered by the accumulation of calcium inside the mitochondria and the increase in oxygen free radicals that accompanies reperfusion.

Although the composition of the MPTP remains to be totally resolved, the prevalent hypothesis is that the pore has three components. These are the adenine nucleotide translocase (ANT), found in the inner of the two mitochondrial membranes, which normally transports ATP out of the

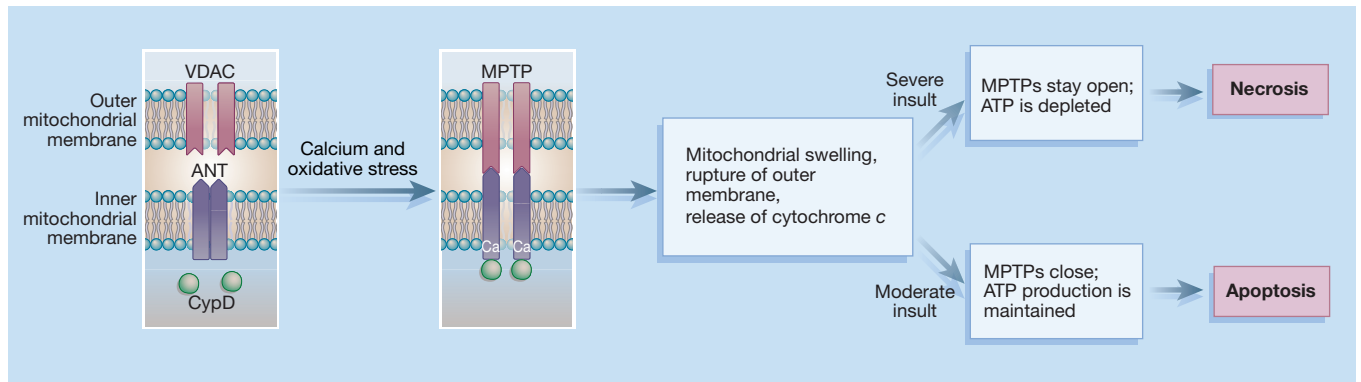


Figure 1 A killing machine. The mitochondrial permeability transition pore (MPTP) is thought to open in response to noxious insults to the cell. Calcium and oxidative stress trigger a conformational change in the adenine nucleotide translocase (ANT) when associated with the voltage-dependent anion channel (VDAC). This is facilitated by bound

cyclophilin D (CypD), and blocked by cyclosporin A, which binds to CypD. MPTP opening leads to mitochondrial swelling, rupture of the outer membrane and release of cytochrome *c*. If pore opening is transient, ATP production is maintained and cells die by apoptosis; but if the pore stays open, ATP is depleted and cells die by necrosis.

mitochondria; the voltage-dependent anion transporter (VDAC) in the outer mitochondrial membrane; and cyclophilin D, a small, soluble protein found inside mitochondria, and so called⁵ because it binds to the immunosuppressant drug cyclosporin A (Fig. 1). This drug prevents cyclophilin D binding to the ANT and strongly inhibits MPTP activity. Cyclophilin D is thought to facilitate a calcium-triggered conformational change in the ANT, converting it to an 'open' pore (Fig. 1).

Although the role of the ANT remains unproven^{6–9}, the work by Nakagawa *et al.*² and Baines *et al.*³ puts the role of cyclophilin D beyond doubt. Both groups investigated mice lacking cyclophilin D ('knockout' mice), and found that the MPTPs in mitochondria from the liver, heart and brain of these mice are no longer sensitive to cyclosporin A; this observation was also reported at the 2005 meeting of the Biophysical Society by Emy Basso *et al.* from the University of Padua, Italy. However, the pore still opens at very high calcium concentrations, just as it does in normal mitochondria treated with cyclosporin A. These results confirm that the role of cyclophilin D is to facilitate a conformational change that can occur in its absence if the stimulus (calcium and oxidative stress) is strong enough.

Hearts from the knockout mice are resistant to reperfusion injury, just as are normal hearts treated with the cyclophilin D inhibitors cyclosporin A and sanglifehrin A (ref. 4). Necrotic cell death induced by oxygen-free radicals or calcium overload was also greatly reduced in liver cells and certain embryonic cells from the knockout mice, again mimicking the effects of cyclosporin A and other MPTP inhibitors. So the original proposal, that the MPTP is key in necrotic cell death caused by calcium overload and oxidative stress¹⁰, is now fully validated.

Perhaps the most important insight provided by the work is that the MPTP does not usually have much of a role in apoptosis.

In many cells, apoptosis is triggered by the release of proteins from the compartment between the two mitochondrial membranes, of which cytochrome *c* is the major player. Once released, this protein activates a cascade of events that prepare the cell for apoptotic death, although the mechanism of cytochrome *c* release remains controversial^{1,11}. One proposed hypothesis is that MPTP opening leads to mitochondrial swelling, rupture of the outer membrane and hence loss of cytochrome *c*. However, many researchers argue that this is unlikely because it would disrupt ATP production, which is required for apoptosis.

The data from the cyclophilin D knockout mice would seem to be the final nail in the coffin of this hypothesis. First, the mice developed apparently normally, so the extensive apoptosis involved in development was not affected by loss of cyclophilin D. Second, by using proteins of the Bcl-2 family that normally stimulate release of cytochrome *c* and apoptosis, both groups found that this process seems to be unaffected in mitochondria and cells isolated from the knockout mice. Third, Nakagawa *et al.*² showed that in all the cell types tested, apoptosis induced by a wide range of stimuli is also unaffected by the absence of cyclophilin D. Together, the two studies lead to an inescapable conclusion: the MPTP is central in necrosis, but not usually in apoptosis.

This same conclusion was reached by Crompton and colleagues¹², who examined a neuronal cell line (B50) synthesizing large amounts of cyclophilin D. They found that the cells were hypersensitive to necrotic cell death induced by calcium and oxidative stress, but were more resistant than usual to apoptosis induced by nitric oxide or staurosporine.

That is not to say that the MPTP is never involved in apoptosis. MPTP opening can be transient, in response to mild insults, and thus does not cause ATP depletion⁴. Should this occur, it might be sufficient to cause

cytochrome *c* release and hence induce apoptosis. Thus, a brief lack of oxygen, or exposure to a toxin that causes limited oxidative stress, may initiate apoptosis (Fig. 1). The advantage of such a mechanism to the tissue would be that it removes damaged cells that are performing inefficiently, but avoids the inflammation that necrosis would bring. Baines *et al.*³ provide support for such MPTP-mediated apoptosis, as they observe an increase in apoptotic heart-muscle cells in normal mice that have high levels of cyclophilin D. These hearts were also enlarged, performed inefficiently and contained swollen, leaky mitochondria. Thus, it seems that, even under normal physiological conditions, excess cyclophilin D leads to sufficient opening of pores to cause cytochrome *c* release and apoptosis.

The results obtained with the cyclophilin D knockout mice pose an intriguing question, however. If such animals live normal, healthy lives, and are protected against injury from oxidative stress, what benefit does cyclophilin D bestow that has led to its conservation through evolution?

Andrew Halestrap is at the Bristol Heart Institute and the Department of Biochemistry, School of Medical Sciences, University of Bristol, Bristol BS8 1TD, UK.

e-mail: A.Halestrap@Bristol.ac.uk

- Green, D. & Kroemer, G. *Science* **305**, 626–629 (2004).
- Nakagawa, T. *et al.* *Nature* **434**, 652–658 (2005).
- Baines, C. P. *et al.* *Nature* **434**, 658–662 (2005).
- Halestrap, A. P., Clarke, S. J. & Javadov, S. A. *Cardiovasc. Res.* **61**, 372–385 (2004).
- Halestrap, A. P., McStay, G. P. & Clarke, S. J. *Biochimie* **84**, 153–166 (2002).
- Bernardi, P., Petronilli, V., DiLisa, F. & Forte, M. *Trends Biochem. Sci.* **26**, 112–117 (2001).
- Waldmeier, P. C. *et al.* *Curr. Med. Chem.* **10**, 1485–1506 (2003).
- Kozoska, J. E. *et al.* *Nature* **427**, 461–465 (2004).
- Halestrap, A. P. *Nature* doi:10.1038/nature02816 (2004).
- Crompton, M., Ellinger, H. & Costi, A. *Biochem. J.* **255**, 357–360 (1988).
- Henry-Mowatt, J., Dive, C., Martinou, J. C. & James, D. *Oncogene* **23**, 2850–2860 (2004).
- Li, Y. M., Johnson, N., Capano, M., Edwards, M. & Crompton, M. *Biochem. J.* **383**, 101–109 (2004).

Thin-film technology

Sticky prospects for coats of diamond dust

J. Am. Chem. Soc. **127**, 3712–3713 (2005)

A diamond coating can make a material robust and resistant to wear, and diamond-coated metal wires have been promoted as super-strong fibres for composites. Yu Liu *et al.* have begun to develop a 'wet chemical' technique for making diamond films under milder conditions than those previously used.

Coating a surface with diamond usually involves chemical-vapour deposition, in which the surface is exposed to carbon-rich vapour at temperatures above 1,000 K in a high vacuum. Liu and colleagues' approach entails chemically 'gluing' diamond nanoparticles to the substrate — in this case, glass — and could ultimately make diamond-coating not just more convenient and economical, but also possible with a wider range of substrates.

Nanoscale diamond powders are commercially available, and Liu *et al.* take advantage of their finding last year that the particles can be coated with fluorine atoms by simply heating them in a mixture of fluorine and hydrogen. The fluorinated diamond can then be covalently attached to the linking groups — amine-tipped alkoxysilanes — on the substrate. The silane group sticks to the glass, and the amine group bonds to the fluorinated diamond, creating a covering of tethered nanoparticles.

Philip Ball

Nuclear physics

Absolutely smashing

Mendeleev Commun. **15**, 1–4 (2005)

Superheavy element 115 of the periodic table, provisionally named ununpentium, was first produced in 2003 by smashing less weighty atoms together. But as the four atoms of the element that were created in the experiment survived for less than a second before decaying, study of the new element was always going to be tricky.

Sergey N. Dmitriev and colleagues seem to have confirmed the existence of element 115 through investigations of its decay chain. Ununpentium was produced by firing a beam of calcium-48 ions at a target made from americium-243. The element then underwent five radioactive decays, each ejecting an α -particle, which created a new isotope of element 105, dubnium-268. Its half-life, about a day, is unusually long for this part of the periodic table.

Dmitriev *et al.* therefore had enough time to separate the collision products from the target, and to isolate dubnium before it underwent spontaneous fission. These

Auditory neuroscience

Sound connection

J. Exp. Biol. **208**, 1209–1217 (2005)

If somebody calls your name across a room, you usually have no trouble working out where they are. This is because the difference between the sound detected by your two ears, in both intensity and time of arrival, provides information about where the sound is coming from.

For lizards it's a different story: their small head size, relative to the wavelengths of many sounds, means that they cannot rely on these principles. Yet experiments carried out by Jakob Christensen-Dalsgaard and Geoffrey A. Manley show, surprisingly, that the lizard ear shows the largest directionality of any land-vertebrate ear studied.

Laser measurements of eardrum motion in four lizard species show that the animals benefit from coupling of the two drums by virtue of a large cavity inside the head. This allows the two sound inputs to interact with each other

and produces up to 50-fold differences between the motion of the two eardrums depending on the sound's direction. The directionality can largely be explained by a simple acoustical model, and would allow the brain to perform a simple subtraction calculation so that the lizard can respond accordingly.

Michael Hopkin



observations provide plausible confirmation that the initial collisions had indeed produced element 115.

Mark Peplow

Cell biology

Tug of war

PLoS Biol. doi:10.1371/journal.pbio.0030100 (2005)

Neisseria gonorrhoeae, the bacterial cause of gonorrhoea, possesses a retractable appendage called a pilus that attaches to cell membranes in infected people. Heather L. Howie *et al.* show that the bacteria use their pili to alter the cells' genetic activity and, in the war with the human host, potentially help the infection to spread.

The team used microarrays to examine the genes that are switched on in human epithelial cells when they are infected with either normal *N. gonorrhoeae* or a strain of the bacterium that can latch on to cell membranes but cannot retract. They identified 52 genes whose activity is boosted by the pili tugging on the cell membrane.

Most of these genes are activated by a signalling protein called mitogen-activated protein kinase (MAPK). The bacterial yanks on the cell membrane seem to trigger MAPK and increase the cell's ability to withstand infection by averting programmed cell death.

The same pathway was activated when Howie *et al.* coated cell membranes with magnetic beads and mechanically tugged on the membranes with a magnet.

The authors speculate that *N. gonorrhoeae* may use the pathway to keep the cells — and infected patients — alive long enough for the bacteria to spread to another host.

Helen Pearson

Optics

The hole story

Opt. Express **13**, 1933–1938 (2005)

Put a pinprick in a sheet of aluminium foil, and light will shine through it. But if the light's wavelength is more than twice as long as the long edge of a rectangular aperture, it is stopped. This 'Rayleigh cut-off condition' seemed to have been contradicted by recent experimental results on light transmission through a rectangular hole in a metal — as the width of the aperture was reduced, the cut-off wavelength actually increased. Reuven Gordon and Alexandre G. Brolo have now developed a mathematical model to explain these effects.

The authors argue that light sets up a resonance motion — known as a 'surface plasmon' — in the electrons along the long edge of the hole. This changes the mode-shape of the light in the hole, leading to a cut-off that increases with decreasing hole width.

For a 15-nm-wide hole, Gordon and Brolo found the maximum transmittable wavelength to be 2.3 times that predicted by the Rayleigh condition for a 'perfect metal'. They therefore advise caution in applying this model, especially as metal holes less than 20 nm across can now be created with nanofabrication techniques.

Mark Peplow

Speed and size of the Sumatra earthquake

We now have a clearer picture of the seismic features of last year's gigantic event.

Our seismological results reveal that Indonesia's devastating Sumatra–Andaman earthquake on 26 December 2004 was 2.5 times larger than initial reports suggested — second only to the 1960 Chilean earthquake in recorded magnitude. They indicate that it slowly released its energy by slip along a 1,200-km fault, generating a long rupture that contributed to the subsequent tsunami. Now that the entire rupture zone has slipped, the strain accumulated from the subduction of the Indian plate beneath the Burma microplate has been released, and there is no immediate danger of a similar tsunami being generated on this part of the plate boundary, although large earthquakes on segments to the south still present a threat.

Our results come from an analysis of the Earth's normal modes ${}_0S_2$, ${}_0S_3$ and ${}_0S_4$. These consist of singlets or split peaks that have distinct periods, or eigenfrequencies, owing to the planet's rotation and ellipticity. Great earthquakes excite these modes, which can be observed by Fourier analysis of long seismograms (Fig. 1a). Singlet amplitudes depend on the location of the earthquake and seismic station, earthquake depth, focal mechanism and seismic moment¹. The decay of energy with time owing to inelastic processes in the Earth, which is equivalent to the width of the spectral peak, depends on the mode's attenuation, or quality factor Q .

Using the focal mechanism and depth reported by the Harvard Centroid-Moment Tensor (CMT) project (see project website, www.seismology.harvard.edu/projects/CMT) and singlet eigenfrequencies², we obtained consistent estimates of seismic moment and Q by two methods: fitting amplitude spectra (Fig. 1a) and fitting the decay of narrow-band filtered singlets³ (results not shown). The estimates of Q for ${}_0S_2$, ${}_0S_3$ and ${}_0S_4$ are 525, 405 and 380, respectively, and are consistent with previously reported values⁴.

As well as the longest-period normal-mode multiplets ${}_0S_2$, ${}_0S_3$ and ${}_0S_4$, we analysed radial modes ${}_0S_0$ and ${}_1S_0$ to obtain estimates of seismic moment and moment magnitude (Fig. 1b). Moment values take into account the inclusion in the seismograms of both ground motion and changes in the gravity field⁵. From the normal modes, we estimate that the seismic moment was as large as 1.0×10^{30} dyn cm (moment magnitude $M_w = 9.3$).

There was also a systematic increase in seismic moment with period (Fig. 1b), which explains why conventional methods used to assess earthquake size dramatically underestimated it. The ${}_0S_2$ moment is about 2.5 times larger than indicated by the CMT solu-

tion, which was based on surface waves with periods below 300 s and which gave a value for the moment magnitude of 9.0. Assuming that other events' reported moments do not also underestimate their true size, this makes the Indonesian earthquake the second largest ever to be instrumentally recorded.

The larger moment we obtain presumably reflects slow slip that was not detectable from the surface waves. The systematic increase in moment with increasing period, reflecting the spectrum of the source time function, is consistent with this idea. A moment still increasing at these very long periods has not been previously observed for other earthquakes, raising issues about the physics of faulting — for example, at what period the moment ultimately stabilizes and reaches its static value.

The slow slip probably occurred over the northern part of the 1,200-km length of the rupture zone indicated by aftershocks (Fig. 1c). The larger moment can be fitted by 11 m of slip on a fault 1,200 km long and 200 km wide (down-dip dimension). This is a larger area than is implied by body-wave inversions, which find rapid slip on the southern part⁶. A larger rupture area is consistent with the fact that split modes are better fitted by a source with centroid at 7°N than by one at the epicentre at 3°N, where rupture started and propagated northward⁶. Another analysis using normal mode ${}_0S_0$ also favours a long rupture⁷. Tsunami run-up, which is the water's highest elevation at the point of maximum horizontal penetration, was 25–30 m in the near field on Sumatra. This implies about 12–15 m of slip, because run-up typically does not exceed twice the fault slip⁸.

It seems that the slow slip helped to excite the tsunami, as suggested by successful modelling of the wave from sea levels detected by the Jason satellite, using a source that includes the northern segment⁹. Large tsunami amplitudes in Sri Lanka and India also support rupture on the northern, north-trending segment, because tsunami amplitudes are largest when perpendicular to the fault.

The picture emerging from the normal modes is consistent with the regional tectonics. Although the plate geometry and motions are not precisely known, the Burma microplate is a sliver between the larger Indian and Sunda plates. Combining the motions of India¹⁰ and Sunda¹¹ with respect to Eurasia, which are known from global-positioning satellite data, with estimates of Burma's motion with respect to Sunda, inferred from back-arc spreading^{12,13}, yields India's motion with respect to Burma (Fig. 1c). Because the India–Burma pole is nearby,

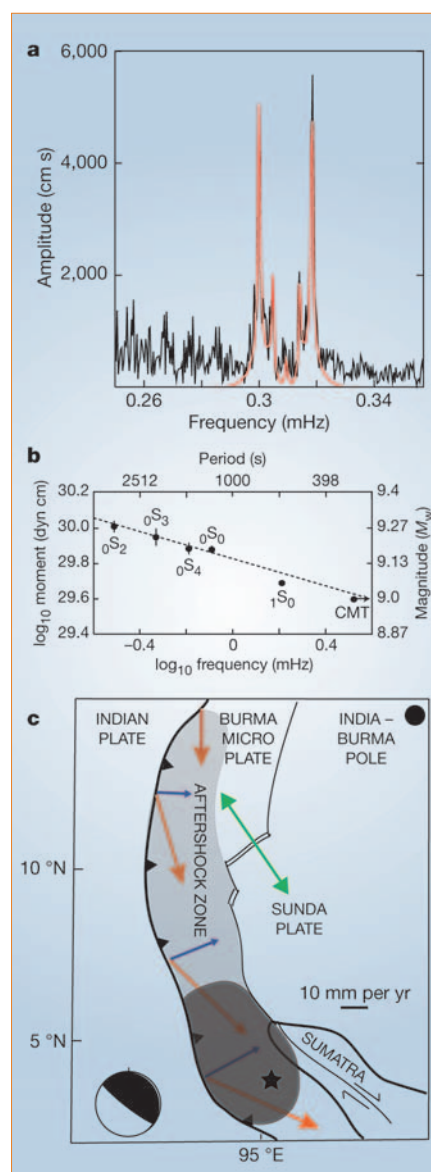


Figure 1 Features of the 2004 Sumatra–Andaman earthquake. **a**, Observed (black) and predicted (red) amplitude spectrum for a ${}_0S_2$ multiplet, showing the best-fitting seismic moment (1.0×10^{30} dyn cm). **b**, Variation in seismic moment and moment magnitude, M_w , with period. CMT (for Centroid-Moment Tensor project) represents the result from surface waves with periods below 300 s. **c**, Comparison of aftershock zone (greys) with minimum area of fast slip (dark grey; corresponding to one-third of rupture area), estimated from body waves, and the possible area of slow slip (light grey; corresponding to the northern part of the fault area) inferred from normal modes. Star, earthquake epicentre. Arrows: total (red) and orthogonal (blue) convergence for an India–Burma euler vector of (14.8°N, 99.8°N) 1.55° per million years; green, back-arc spreading; scale bar, 10 mm per year. Black and white disc, CMT focal mechanism.

the convergence direction varies along the rupture zone and becomes strike–slip at the north end of the rupture, presumably

explaining why rupture ceased. The CMT focal mechanism reflects the arc-normal component of convergence: 15–25 mm per yr.

If the entire aftershock zone slipped, then strain accumulated on the northern part of the rupture has been released. There is therefore no immediate threat of an oceanwide tsunami being generated by slip on this segment of the plate boundary, because such earthquakes should be at least 400 years apart. However, the danger of a large tsunami resulting from a great earthquake on segments to the south remains.

Seth Stein, Emile A. Okal

Department of Geological Sciences, Northwestern University, Evanston, Illinois 60208, USA
e-mail: seth@earth.northwestern.edu

1. Stein, S. & Geller, R. J. *Phys. Earth* **25**, 117–142 (1977).
2. Dahlen, F. A. & Sailor, R. V. *Geophys. J.* **58**, 609–624 (1979).
3. Geller, R. & Stein, S. *Bull. Seismol. Soc. Am.* **69**, 1671–1691 (1979).
4. Stein, S. & Nunn, J. *Bull. Seismol. Soc. Am.* **71**, 1031–1047 (1981).
5. Dahlen, F. A. & Tromp, J. *Theoretical Global Seismology* (Princeton Univ. Press, Princeton, New Jersey, 1998).
6. Ji, C. http://neic.usgs.gov/neis/eq_depot/2004/eq_041226/neic_slav_ff.html
7. Park, J. <http://www.iris.edu/sumatra>
8. Okal, E. A. & Synolakis, C. E. *Geophys. J. Int.* **158**, 899–912 (2004).
9. Pacific Marine Environmental Laboratory <http://www.pmel.noaa.gov/tsunami/research.html>
10. Sella, G. E., Dixon, T. H. & Mao, A. J. *Geophys. Res.* **107**, doi:10.1029/2000JB000033 (2002).
11. Chamot-Rooke, N. & Le Pichon, X. *Earth Planet. Sci. Lett.* **173**, 439–455 (1999).
12. Curry, J. R. et al. in *Geological and Geophysical Investigations of Continental Margins* (eds Watkins, J. S. et al.) 189–198 (Am. Assoc. Pet. Geol., Tulsa, Oklahoma, 1979).
13. Bird, P. *Geochim. Geophys. Geosyst.* **4**, doi:10.1029/2001GC000252 (2003).

Competing financial interests: declared none.

Seismology

Energy radiation from the Sumatra earthquake

We determined the duration of high-frequency energy radiation from Indonesia's great Sumatra–Andaman earthquake (26 December 2004) to be about 500 seconds. This duration can be translated into a rupture length of about 1,200 km, which is more than twice as long as that inferred from body-wave analyses performed soon after the event. Our analysis was able rapidly to define the extent of rupture, thereby aiding the assessment of seismic hazard in the immediate future.

Soon after the Sumatra–Andaman earthquake, seismic body-wave studies in the period range of 10 to 50 s indicated that there had been a slip distribution over a 400-km segment^{1–3} (Fig. 1a). These methods for rapid assessment of major earthquakes rely on the extended P-wave train to deduce the source rupture pattern. But when an event lasts longer than the period between the later-phase PP and P waves, a problem arises in determining the source duration.

Figure 1b shows the observed displacement seismogram for a seismic event of magnitude 7.1 (which occurred near the

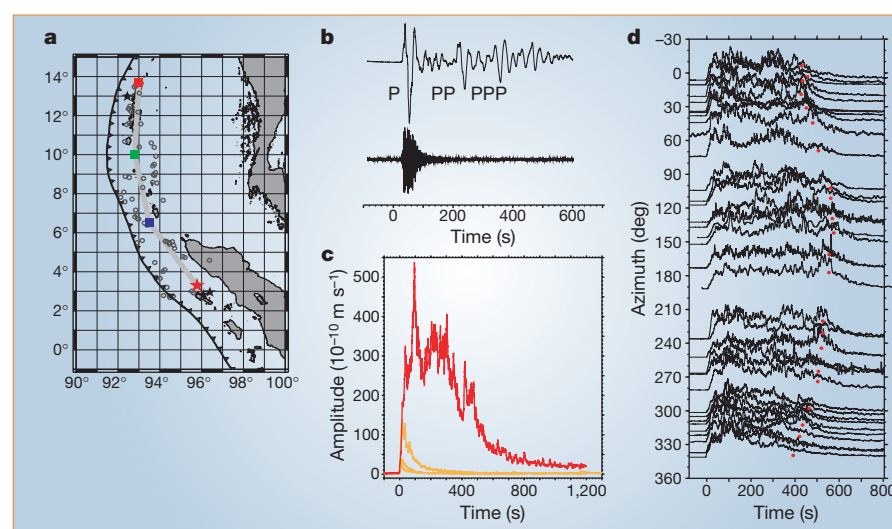


Figure 1 Frequency of radiation from the 2004 Sumatra–Andaman earthquake. **a**, Rupture-termination points of the earthquake estimated from body-wave inversion (blue square) and from high-frequency radiation (red square) calculations; green square, candidate for termination point (see supplementary information). Red star, earthquake epicentre; circles, aftershock locations; black stars, locations of large foreshocks and aftershocks. **b**, Typical teleseismic seismograms (broadband) showing P waves before (top) and after (bottom) high-bandpass (2–4 Hz) filtering; later phases are removed by attenuation. $D=64$ degrees. **c**, Enveloped high-frequency seismogram comparing the main shock (red) with smaller events (orange) at the same station. **d**, Smoothed envelopes (2–4-Hz bandpassed) of the main shock as a function of azimuth (horizontal angle): shorter wave trains are evident in the direction of rupture (azimuth about 340°). Red dots, estimated end of the source duration.

epicentre of the 26 December earthquake (Fig. 1a, red star) on 2 November 2002). The phases PP and PPP are apparent, but are seen to disappear at high frequency (in the range 2 to 4 Hz) because of attenuation in the Earth's mantle⁴. We therefore analysed high-frequency radiation from the Sumatra earthquake by determining amplitude–time envelopes and smoothing them as described⁴ (see supplementary information). Comparison of the smoothed envelopes for the main shock of this event with three smaller fore- and aftershocks shows that the amplitude and duration of the main shock are much larger (Fig. 1c).

Figure 1d plots the envelopes for the main shock as a function of azimuth (angular distance from the horizon). Each envelope shows a short rise time, then a relatively flat, sustained portion, which is followed by rapid decay. The envelope duration (red dots in Fig. 1d) reveals a clear azimuthal pattern of 400 s in the direction of rupture (at about 340°) to 590 s in the opposite direction. We assume that these high-frequency signals are derived mostly from the rupture front⁵. From the range of the azimuthal variation in duration (190 s), we determine the rupture length to be 1,200 km — the longest ever recorded. Then, from the average duration (about 500 s), we can derive the average rupture speed as 2.5 km s^{-1} .

The rupture length is comparable to the length of the aftershock distribution (Fig. 1a). The rupture was substantially longer than that of the 1960 great Chilean earthquake⁶ (340 s), and had a larger magnitude of 9.5, but the tectonic setting at Sumatra is very different from that at Chile. In Chile, the age

of the subducting plate is young (15 million years old) and the plate convergence is nearly normal to the trench, whereas in Sumatra the subducting plate is older (more than 60 million years old) and the plate convergence is oblique, especially in the north. This difference could be responsible for the difference in slip behaviour between the two events, and hence the disparities in rupture length and magnitude.

The high-frequency radiation reflects only the propagation of the rupture front, so it alone cannot uniquely determine the slip distribution. Modelling of long-period surface waves and normal modes will eventually constrain the spatiotemporal distribution of slip. Despite such uncertainty, our simple analysis at high frequency provides an accurate and rapid determination of the duration and rupture length of this earthquake, which are important for rapid assessment of immediate seismic hazard in the area.

Sidao Ni*†, **Hiroo Kanamori**†, **Don Helmberger**†

*School of Earth and Space Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China
e-mail: sdni@ustc.edu.cn

†Tectonics Observatory, California Institute of Technology, Pasadena, California 91125, USA

1. Ji, C. www.gps.caltech.edu/~jichen
2. Yamanaka, Y. www.eri.u-tokyo.ac.jp/sanchu/Seismo_Note/
3. Yagi, Y. <http://iisec.kenken.go.jp/staff/yagi/eq/Sumatra2004/Sumatra2004.html>
4. Shearer, P. M. & Earle, P. S. *Geophys. J. Int.* **158**, 1103–1117 (2004).
5. Madariaga, R. *Ann. Geophys.* **1**, 17–23 (1983).
6. Houston, H. & Kanamori, H. *Bull. Seismol. Soc. Am.* **76**, 19–42 (1986).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Chromatin remodelling and epigenetic features of germ cells

Sarah Kimmins & Paolo Sassone-Corsi

Institut de Génétique et de Biologie Moléculaire et Cellulaire, B.P. 10142, 67404 Illkirch, Strasbourg, France

Germ cells have the unique capacity to start a new life upon fertilization. They are generated during a sex-specific differentiation programme called gametogenesis. Maturation of germ cells is characterized by an impressive degree of cellular restructuring and gene regulation that involves remarkable genomic reorganization. These events are finely tuned, but are also susceptible to the introduction of various types of error. Because stable genetic transmission to future generations is essential for life, understanding the control of these processes has far-reaching implications for human health and reproduction.

Cells transmit information to the next generation via two distinct routes, genetic and epigenetic. While genetic inheritance is based on the DNA code, epigenetic information comprises modifications occurring directly on DNA or on the chromatin, a proteinaceous complex associated with DNA. Whereas the major type of DNA modification is the methylation at cytosines, there are multiple modifications associated with chromatin. Their inheritability has been demonstrated only in rare cases^{1,2}.

The basic unit of chromatin, the nucleosome, is composed of a histone octamer that includes the histones H2A, H2B, H3 and H4. Adjacent nucleosomes are connected by histones of the H1 linker class. Various covalent modifications occur on specific residues of histones, including methylation, phosphorylation, acetylation and ubiquitination. Combinations of these are thought to contribute in various ways to chromatin organization and gene expression. Several excellent review articles on the molecular mechanisms of epigenetic control constitute useful further reading^{3–6}.

In the lifetime of a mammal, two periods are characterized by epigenetic reprogramming—gametogenesis and early embryonic development. During spermatogenesis, global sex-specific changes to the epigenome occur as a wave of DNA demethylation, followed by DNA methylation and chromatin modifications^{7,8}. This may contribute to a unique feature of the germ line—the control of gene function from one generation of an organism to the next.

The epigenetic control of gene function in germ cells follows highly specialized programmes. In particular, there are striking sex-specific differences in the development of male and female gametes, and in the epigenetic state of their genome^{7,8}. During gametogenesis the male- and female-specific epigenetic programmes are ‘reset’, a remodelling of the epigenome, to allow the gametes to fuse at fertilization and give rise to a zygote that is totipotent and thus able to give rise to any cell type. The fidelity of the resetting process is important for preventing aberrant epigenetic modifications that can be passed on to the next generation. Hence, the implications for human health are profound. Male and female germ cells from parents who had a failure in epigenetic reprogramming—from either undergoing assisted reproductive manipulations⁹ or exposure to harmful environmental or chemical factors—produce offspring with a greater susceptibility to disease¹⁰.

Germ cells differ from somatic cells in that they exhibit a number of specialized regulatory pathways, either by expressing specific transcription factors or isoforms. These might form unique regulatory complexes that may involve specific chromatin components unique to germ cells. Does the highly dynamic, sex-specific nature of chromatin remodelling in germ cells contribute to these differences?

Indeed, germ cells possess a remarkably diverse set of histone

variants¹¹—highly abundant proteins thought to be important for chromatin organization. One interesting sex-specific difference is that in the male many variants are expressed in a highly regulated temporal manner, underscoring that spermatogenesis has evolved in a different fashion from oogenesis. In addition, maturing sperm cells undergo an extraordinary process of chromatin remodelling called the histone-to-protamine transition, which reshapes the nucleus and compacts chromatin in an unparalleled architecture. These events prepare germ cells for fertilization and thereby constitute essential regulatory processes. In this review we bring into focus the molecular players and regulatory pathways involved in chromatin remodelling and gene regulation in germ cells. The epigenetic programme of male germ cells is especially emphasized because it shows characteristics that are highly distinct from somatic cells.

Sex-specific routes to haploidy

Sex-specific differences in the development and cellular organization of male and female gametes begin as early as meiosis, the reductive cell division through which haploidy is achieved. In the male, early spermatocytes undergo the S (synthesis) phase of the cell cycle, giving rise to tetraploid spermatocytes, marking the beginning of meiotic prophase. At this point, the homologous chromosomes pair up and become joined by a proteinaceous scaffold called the synaptonemal complex¹². At this time, genetic recombination occurs and paired chromosomes align on the metaphase plate for segregation of sister chromatids into two daughter cells.

After meiosis, spermatogenesis follows a tightly regulated programme during which, under the control of the hypothalamic–pituitary–gonadal axis, major morphological and biochemical changes occur, including cytoplasm elimination and nuclear reshaping. Chromatin remodelling in late spermiogenesis is particularly impressive. During this process, most somatic histones are replaced by DNA packaging proteins that are unique to male germ cells—namely the transition proteins, which are later replaced by the protamines. The incorporation of protamines into sperm chromatin induces DNA compaction, and is followed by cytoplasmic ejection, and acrosome and flagellar formation⁸.

The female gamete needs to produce factors that are involved in metabolism and early development and therefore, in contrast to the sperm, which ejects most of its cytoplasm, has evolved an enriched cytoplasm. At the first cellular division of the primary oocyte the metaphase spindle is polarized at one end of the cell, resulting in the production of two daughter cells, one large with much cytoplasm (secondary oocyte), and the other with very little cytoplasm (the polar body) (Box 1). In meiosis II, similar mechanisms operate to ensure that the mature egg conserves a large cytoplasm¹².

Unique chromatin architecture in male germ cells

Spermatogenesis is characterized by a particularly spectacular chromatin remodelling process, in which somatic linker histones are sequentially replaced by testis-specific variants, followed by the replacement of most histones with protamines. Are the germ cell-specific histones and special chromatin organizing proteins that are required for these events unique to meiosis? Are they involved in the ability of sperm to fertilize the egg and do they thereby give rise to embryos? In this section we explore the unusual chromatin composition of the mammalian male germ cell and consider its role in sperm development and function.

The transition proteins

Transition proteins are thought to prepare the chromatin for association with protamines, possibly by influencing DNA condensation¹³. During development of the post-meiotic haploid spermatids, transition proteins become a significant chromatin component. After histone removal and before protamine deposition, transition proteins constitute 90% of all chromatin basic proteins. The best characterized of these proteins, transition protein 1 and transition protein 2, represent about 55% and 40% of spermatid total nuclear proteins, respectively¹³. Transition protein 1 is a small basic protein of 54 residues, rich in arginine, lysine and serine. Twice the size of transition protein 1, transition protein 2 is enriched in basic residues in its carboxy terminus and contains two putative zinc fingers in the amino-terminal region. Mice mutants for transition proteins 1 and 2 are able to produce offspring, although with reduced fertility, suggesting overlapping roles of these proteins^{14,15}.

What are the signals that trigger the incorporation of transition proteins into chromatin? How is the timing of this event coordinated with histone replacement? Transition proteins become phosphorylated in their basic domain at synthesis, followed by dephosphorylation that seems to facilitate DNA binding and

chromatin condensation. In transition protein 2, both Ser 109 and Thr 101 are potential phosphoacceptor sites for the cAMP-dependent PKA (protein kinase A)¹⁶. Notably, phosphorylated transition protein 2 is associated with less condensed DNA, which might facilitate protamine entry¹⁶. It is believed that histone H4 hyperacetylation facilitates the transition to protamines¹⁷. While it is unclear how H4 hyperacetylation may elicit this function, it is notable that CDY, a Y-chromosome encoded histone acetyltransferase protein, is testis-specific and readily acetylates H4 (ref. 18).

The protamines

Protamines are small proteins (relative molecular mass 4,000–12,000) that are evolutionarily related to histone H1 (refs 19, 20), but have significantly different biochemical properties. Somatic histone H1s have lysine-rich N- and C-terminal tails, and very low arginine content. In contrast, protamines have very low lysine content, and more than 50% of their residues are arginine, which is probably responsible for their high DNA-binding affinity. This can be attributed to the fact that arginine has a greater flexibility in the formation of hydrogen bonds with the DNA backbone owing to its complex guanidinium group²¹. It seems likely that during the evolution of protamines from H1 histone, protamines acquired higher arginine content, which may have resulted in their improved chromatin condensing properties.

Most mammals express only protamine 1, whereas mice and humans express two protamines. Their structures differ and disruption of either gene in the mouse results in male infertility²². Similar to transition proteins, protamine phosphorylation seems to lead to incorporation into DNA¹⁹. Protamine 2 undergoes phosphorylation by Ca^{2+} /calmodulin-dependent protein kinase IV (CamK4), and disruption of the *CamK4* gene in the mouse results in failure to exchange transition protein 2 for protamine 2, and subsequently blocks germ cell differentiation²³. However, CamK4 may not be solely responsible. Another candidate kinase is the

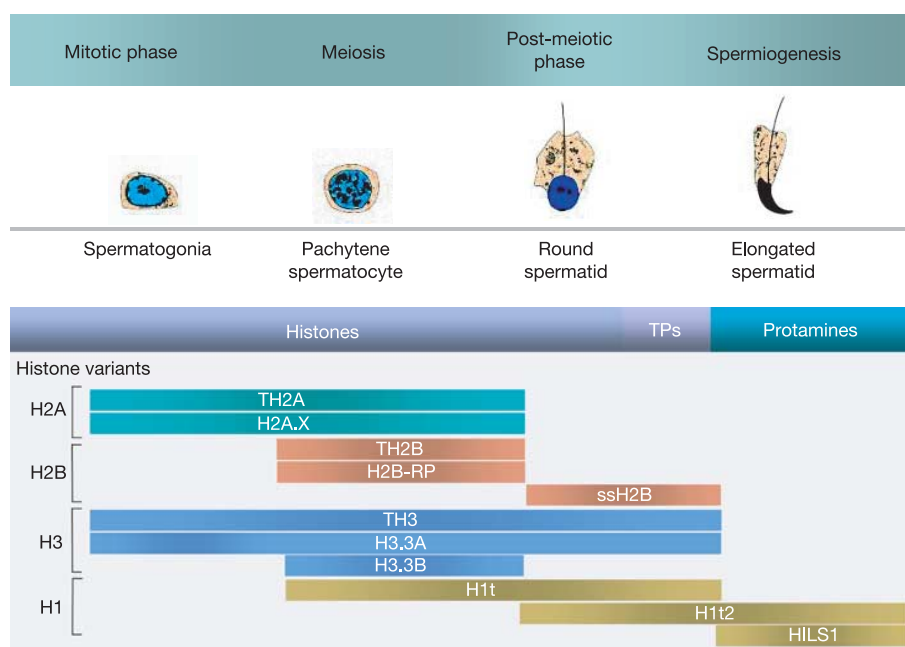


Figure 1 Unique chromatin remodelling during the development of male germ cells. Spermatogonia generate preleptotene spermatocytes, which enter meiosis. Next, round spermatids enter the spermiogenic phase. Germ cells contain many histone variants, which are expressed at different stages of male germ cell development. During the last phase of spermiogenesis, chromatin is highly compacted, a process

that includes the replacement of most histones with sperm-specific transition proteins (TPs), which are subsequently replaced by protamines. Coincident with chromatin remodelling and reshaping of the sperm head is the formation of acrosome and tail. For details of histones ssH2B see ref. 78, and for H2B-RP see GenBank accession number AAP57490.

cAMP- and Ca^{2+} -independent casein kinase II (CK2). The CK2 α' catalytic subunit isoform is preferentially expressed in male germ cells and its ablation in the mouse results in spermatogenic defects, including morphologically abnormal sperm²⁴. Whatever the signalling pathways involved, the timing of the histone-to-protamine transition is a finely tuned event. Indeed, premature expression of protamine 1 in transgenic mice leads to precocious chromatin condensation²⁵.

A multitude of histone variants

Male germ cells have an unusually high number of histone variants in comparison to somatic cells (Figs 1–3). There are two types of variants: those like the testis-specific histone H4, with minor or no amino-acid differences from the somatic H4 (ref. 26); and variants with different amino-acid sequences and structure, such as H2A.X (ref. 27). Several characteristics distinguish variants from their somatic counterparts. First, their messenger RNA synthesis is often uncoupled from DNA replication, and accordingly they lack a stem-loop in the transcript 3' end, a structure classically required for cell-cycle-regulated degradation²⁸. Instead, transcripts encoding variants carry longer poly-A tails, which increases their stability²⁹. Second, histone variant genes are not organized in clusters, but are solitary and may contain introns³⁰.

Incorporation of variants into the nucleosome could influence

gene regulation. Indeed, H2A.Z and H3.3 participate in distinct nucleosome assembly pathways, possibly specific for DNA-synthesis-independent chromatin deposition. H2A.Z is deposited by the SWR1 complex³¹, whereas H3.3 is deposited by the HIRA complex³². This specificity implies that 'variant' nucleosomal octamers could be deposited in a highly regulated fashion, so that their position would impose inheritable boundaries of active or inactive chromatin.

Histone H3 variants

Germ-cell-specific histones appear in spermatogonia³³, and later in spermatids^{34,35}, but the majority are synthesized and incorporated into chromatin during meiosis (Fig. 1). H3.3A is an interesting variant that is incorporated into the germ line, and is present during phases of development from spermatogonia to spermatid³³. By prophase I, histone H3 is largely replaced by H3.3A and H3.3B, which apparently associate with euchromatin^{33,36}, suggesting a role in the massive transcription programme in spermatocytes⁸. Indeed, H3.3 is relatively enriched in modifications associated with transcriptional activation and is deficient in dimethyl Lys 9 (ref. 37), a modification associated with silencing and heterochromatin formation³. Finally, as H3 phosphorylation at Ser 10 is coupled to mitotic chromosome condensation³⁸, it is appealing to speculate a role for phosphorylated H3 variants at meiosis.

Of all the histone variants, CENP-A (centromere protein A) possesses features suggestive of a critically important role in epigenetic function. Located at the centromere, it is a divergent paralogue of H3, with which it shares very little homology in the N-terminal tail (Fig. 2). CENP-A is not displaced during the histone-to-protamine transition and thus behaves as an inherited element, conserved in the two daughter chromatids during S phase. Interestingly, clear differences can be spotted in N-terminal tails between somatic histones and germ-cell-specific variants (Fig. 2). Despite the divergences with H3, human CENP-A undergoes phosphorylation by Aurora-B at Ser 7 (ref. 39), a residue absent in mouse CENP-A, which may instead become phosphorylated at Ser 15. The inheritability of CENP-A hints at tantalizing roles in fertilization.

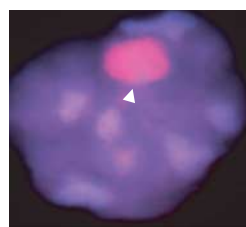
Histone H2A and H2B variants

The H2A and H2B variants possess notable differences in their N-terminal tails. For example TH2B, a testis-specific variant of H2B (ref. 40), differs by the addition of three potential phosphorylation sites (Ser 12, Thr 23 and Thr 34) and the repositioning of two others (Ser 5 and Ser 6), resulting in a different 'phosphorylation map' of the N-terminal tail (Fig. 2). Germ cell kinases, such as testis-specific isoforms of protein kinase C⁴¹ and Aurora-C⁴², could potentially target TH2B through these sites. In addition, in the context of the binary switch hypothesis of combined histone modifications³, the insertion of specific phosphoacceptor sites in TH2B generates combinatorial associations of lysine, serine and threonine residues, which could impart unique patterns of acetylation and/or methylation (Fig. 2). TH2B nuclear distribution seems unequal, suggestive of association with specific chromatin domains⁴⁰.

In H3.3, the Thr-Lys and Lys-Ser binary sites are perfectly conserved with respect to H3, although an Ala > Ser change (Ser 31) in H3.3 generates a putative phosphoacceptor site for cell-cycle-regulated kinases (Fig. 2). In H3.3, a specialized distribution of methylation versus acetylation modification is indicative of association with transcriptionally active chromatin³⁷, and suggests that it functions in post-meiotic gene expression.

The case of H2A.X is particularly interesting. A single Glu > Thr change (Thr 7) generates a potential binary site that might be targeted by a testis-specific kinase. In addition, the Ser 139 phosphoacceptor site in H2A.X is a target for kinases of the PI3 family, including DNA-PK⁴³, ATM (ataxia telangiectasia mutated)⁴⁴ and the ATM-related protein, ATR⁴⁵. H2A.X is implicated in meiosis and is thought to be involved in chiasmata formation because of its

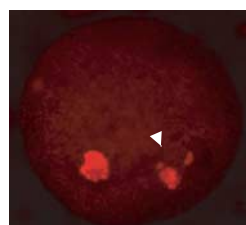
Box 1



Sex (XY) body

The sex chromosomes of mammalian spermatocytes form a specialized nuclear territory known as the XY (or sex) body, where both transcription and homologous recombination are restricted. Proteins assembled into the XY body are typical of heterochromatin. Here the XY body is marked in a

pachytene spermatocyte using an antibody against methylated histone H3-Lys 9 (arrowhead).



Polar body

The polar body is a minute structure that separates by karyokinesis from the ovum during its maturation. In the maturation of ordinary ova two polar bodies are formed. The first is usually larger than the second one, and often divides into two after its separation from the ovum. Each of the polar

bodies removes maternal chromatin from the ovum to make room for the chromatin of the fertilizing spermatozoa. The polar body exhibits phosphorylation of histone H3 on Ser 10 (P-H3), a hallmark of chromosome condensation during mitosis and meiosis. Pictured is an oocyte collected 16 h after ovulation and immunolabelled for P-H3 (arrowhead).



Chromatoid body

The chromatoid body is a germ cell-specific cytoplasmic organelle that closely interacts with the nucleus and contains compacted mRNA. Its localization is shown in a spermatid that has been immunolabelled for the mouse VASA homologue, MVH, an RNA helicase (arrowhead). Some

intriguing observations suggest that the chromatoid body could function to monitor mRNA localization.

function at sites of double-strand-breakage⁴⁶. Importantly, H2A.X deficiency results in arrested spermatogenesis at the pachytene stage⁴⁶, a phenotype reminiscent of the ATM-null mice, which show severe meiotic disruption⁴⁷. ATM and ATR kinases are involved in the cellular response to ionizing radiation and DNA double-strand break-inducing events. ATM also targets the cohesin SMC1, p53 and the checkpoint kinase Chk2 (ref. 48), stressing its implicated role in meiotic chromosome dynamics. ATM is present in the spermatocyte cellular compartment, known as the XY body or sex body, where the X and Y chromosomes are unsynapsed (Box 1).

The sex body

The sex body is a nuclear compartment in germ cells where RNA polymerase II is absent⁴⁹, and where selective inactivation of the sex chromosomes occurs. In male meiosis, the heteromorphic X and Y chromosomes undergo the condensation process of heterochromatinization, accompanied by transcriptional silencing. It is believed that the sex body excludes promiscuous pairing or recombination between nonhomologous chromosomes, thereby reducing the risk of aneuploidy⁴⁹.

The sex body contains a number of proteins that are implicated in heterochromatinization, including the H2A histone variant, macro-H2A1, and heterochromatin protein 1 (HP1 β)⁵⁰. Both are known to be involved in meiotic sex chromosome inactivation (MSCI) and formation of the XY body. Another feature of the sex body is ubiquitination of H2A, which suggests it has a role in gene silencing⁵¹.

The special case of the histone H1 family

H1 linker histones influence the degree to which chromatin folds, by virtue of their association with the DNA-connecting nucleosomes. Proteins of the H1 family greatly diverge in their structure, and several are germ-cell-specific variants (Fig. 3). The C termini of the variants are highly different, which may influence the degree to which they bind chromatin⁵². Additional α -helices are predicted to

form in the C terminus of H1t2, HILS1 and H1Foo (Fig. 3), which could influence interaction with nuclear components.

The two testis-specific H1 variants, H1t (ref. 53) and H1t2 (ref. 35), show a high degree of identity. The restructuring of sperm chromatin begins during meiotic prophase when the somatic linker histones, H1A and H1B, are displaced by the variant H1t (ref. 54). Targeted deletion of H1t or H1.1 does not affect fertility, suggesting functional redundancy^{53,55}.

Another testis-specific linker histone is HILS1 (histone H1-like protein in spermatids 1), which is restricted to elongating spermatids⁵⁶. The nuclear distribution of HILS1, transition protein 2 and protamine 1 (ref. 56) is identical, suggesting that HILS1 is intimately linked to chromatin condensation—a process that may involve histone kinases³ for which there are multiple potential sites in HILS1. Importantly, of the five Ser/Thr cyclin/CDK sites in somatic histone H1 that seem to be important for dynamic nuclear mobility, several are absent in the variants, particularly H1t and H1t2 (ref. 57).

Methylation of somatic H1 by the lysine methyltransferase Ezh2 (human Enhancer of Zeste homologue) at Lys 26 seems to be important for transcriptional repression⁵⁸. This site is markedly conserved among somatic H1 isoforms and shares a striking similarity to the Lys 27 site in histone H3, which is targeted by a Ezh2-containing complex with a different histone lysine methyltransferase specificity. Testis H1 variants have glycine or alanine residues in the place of Lys 26, which makes them unlikely Ezh2 targets, and suggests that testis-specific H1 variants are engaged in more dynamic and regulatory functions in germ cells.

The recently described H1t2 seems to play a highly specialized role in establishing cell polarity and, similar to protamines, in directing chromatin condensation³⁵. H1t2 appears in round spermatids, in a territory underneath the nuclear membrane and basal to the presumptive acrosome, where chromatin condensation initiates. H1t2 may be a component of the chromatin organizing centre, as indicated by the impaired chromatin condensation in the sperm nucleus of H1t2-null mice³⁵. Like other variants, the H1t2 C terminus is enriched in putative phosphorylation sites (Fig. 3). H1t2 seems to be as important as protamines in chromatin reorganization and constitutes a notable example of a histone with highly selective intranuclear distribution.

The female side of chromatin remodelling

In the oocyte, acetylation is the most dominant form of histone modification. The different patterns of lysine acetylation observed on histones H3 and H4 in meiotic oocytes and preimplantation embryos, suggests that acetylation is important in epigenetic reprogramming and possibly in chromosome dynamics. H3 and H4 lysines (with the exception of H4 Lys 5) are deacetylated during meiosis in mouse oocytes, as well as in somatic nuclei that have been transferred into enucleated oocytes⁵⁹. The global reduction in acetylation levels that has been observed in maturing oocytes is thought to help reprogramming by erasing information on active genes. Indeed, histones H3 and H4 are highly acetylated in regions of active gene expression and underacetylated on silent genes².

Furthermore, histone deacetylation in meiotic oocytes facilitates the binding of ATRX, a member of the SWI/SNF helicase family of chromatin-modifying proteins⁶⁰. ATRX localizes to centromeric heterochromatin, and alignment of chromosomes on the metaphase plate is abnormal when ATRX's function is blocked⁶⁰. As ATRX is dependent upon histone deacetylation, removal of this mark seems critical for oocyte differentiation and chromosome segregation. In contrast to the rapid deacetylation occurring at germinal vesicle breakdown, histone methylation remains throughout meiosis and is probably required for HP1 binding to the centromeric domain⁶⁰.

During oocyte development, chromatin composition is altered by the replacement of somatic histone H1 with a specialized maternal



Figure 2 Modifications of histone variants. N-terminal tails of generic histones and some testis-specific variants, with red residues indicating differences. The indicated modifications (red, phosphorylation; green, acetylation; blue, methylation) are either demonstrated^{3,5} (filled circles) or putative (open circles). Shaded boxes indicate binary-switch sites (blue for Thr/Lys associations, red with Ser). Various differences exist in H2A.X and TH2B. The Ser 31 change in H3.3 generates a putative phosphorylation site. The N-terminal tail of CENP-A contains several putative phosphorylation sites (Ser 7 is demonstrated as such). There are nine prolines (H3 has only two) and one lysine as putative sites for modifications (H3 has eight lysines).

variant, H1Foo (H1 histone family, oocyte-specific; formerly named H1oo). H1Foo is distinct from its various male counterparts in that its mRNA is heavily polyadenylated and the protein itself is highly rich in lysines, which are potential sites for methylation or acetylation. The N terminus contains ten potential phosphorylation sites, whereas the extended C terminus (Fig. 3a) could facilitate chromatin condensation or influence mobility. H1 subtypes are known to play a role in gene silencing in *Caenorhabditis elegans*⁶¹, so H1Foo may serve a similar purpose in mammals. H1Foo appears first in secondary follicles and disappears by the four-cell embryonic stage in the mouse⁶². Transcription decreases during oocyte development, as the chromosomes condense⁶³. It is possible that H1Foo acts as a transcriptional repressor by altering chromatin structure. There is correlative data to suggest this is the case—H1Foo is present during the time that oocyte transcription is low, and is rapidly removed from the chromatin when zygotic gene activation occurs⁶⁴.

Thus, whereas male germ cells incorporate a multitude of H1, H2A, H2B and H3 variants—and undergo extraordinary chromatin remodelling—the oocyte maintains a chromatin structure much

like somatic cells. In this scenario, the incorporation of H1Foo could constitute an important exception because it may help define the active/inactive transcription borders that are likely to be established upon fertilization.

Specialized transcription rules in male germ cells

The impressive wave of transcription that occurs post-meiotically in male germ cells is achieved through assembly of transcription complexes and specialized chromatin remodelling enzymes. Studies on the nuclear factors CREM, TLF, TAF7L and related co-activators, have revealed a number of interesting germ-line-specific features discussed elsewhere⁸. For example, CREM-dependent transcriptional activation in somatic cells requires CREM phosphorylation and subsequent CBP recruiting. In contrast, in germ cells the requirements of phosphorylation and histone acetyltransferase function⁸ is bypassed by a specific coactivator ACT (activator of CREM in testis).

Transcriptional regulation also plays an important role early in spermatogenesis, when spermatogonial cells face the choice of self-

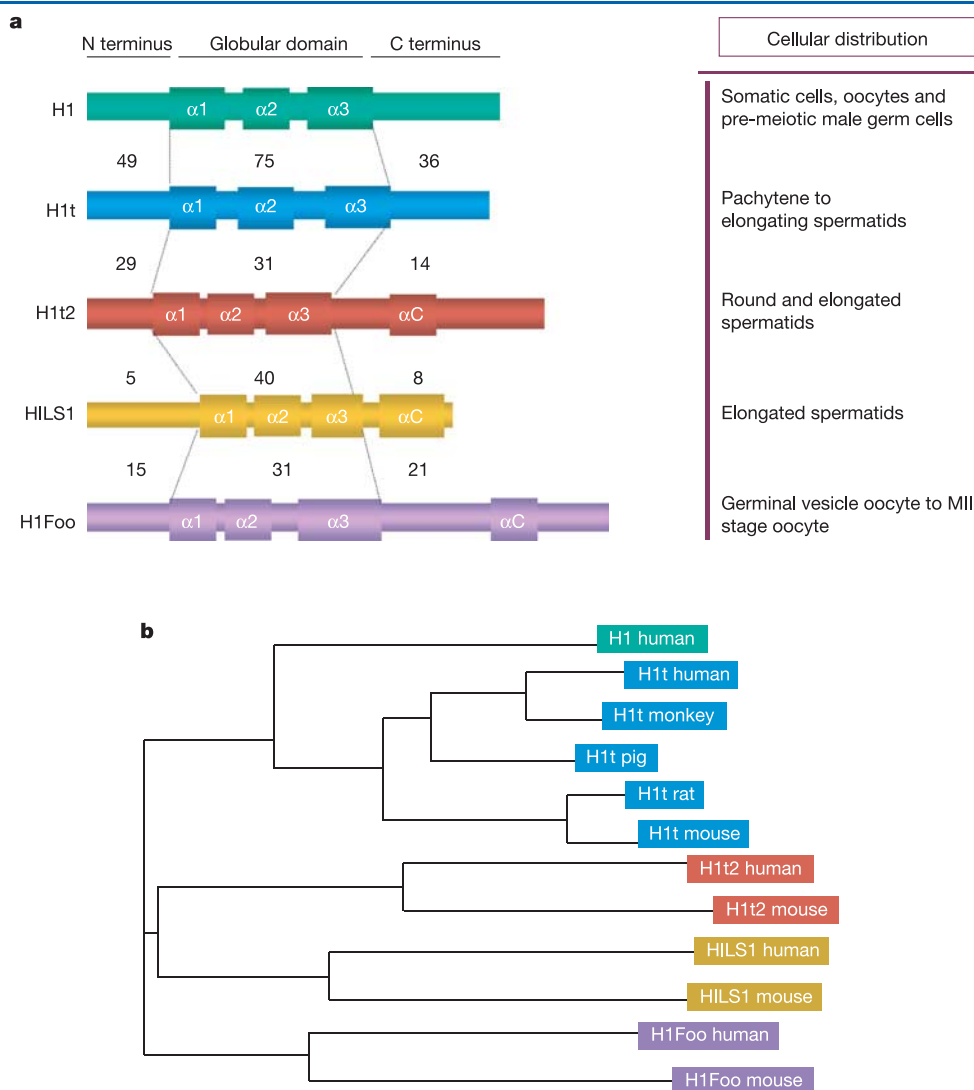


Figure 3 Germ cells contain various H1 variants. **a**, Distribution and secondary structure of generic human histone H1 and germ cell-specific H1 variants. Numbers represent percentage identity against generic histone H1, as determined by the EMBOSS-align algorithm (www.ebi.ac.uk/Tools/sequence.html). H1 proteins have a multidomain structure with a central relatively conserved globular domain containing

three helical regions flanked by variable N- and C-terminal domains. The region of the third α -helix in H1t2 is highly divergent. H1LS1 secondary structure is based on the predicted structure⁵⁴. **b**, Phylogram of the H1 family. The most similar to H1 is H1t, and the least similar is the oocyte-specific H1Foo. Altered C termini are notable as they may modulate interaction with chromatin.

renewal or differentiation. One of the players in this decision-making process is Plzf (promyelocytic leukaemia zinc finger), a nuclear factor of the POK (POZ and Krüppel) family of transcriptional repressors. The POZ domain of Plzf recruits Polycomb proteins such as BMI1, which subsequently recruits histone deacetylases. The role of BMI1 and other Polycomb proteins is to maintain stable and heritable repression of specific developmental genes; thus it is possible that through this mechanism, Plzf-dependent histone deacetylases regulate chromatin remodelling in defining spermatogonial cell fate. Importantly, Plzf-null mice show progressive loss of spermatogonial cells and increased apoptosis^{65,66}.

Because DNA methylation plays a central role in male germ cell differentiation, its potential influence on histone methyltransferases may have far-reaching consequences. Indeed, there are several histone methyltransferases, which show hallmarks of being important regulators of germ cell transcription. One is the mammalian H3 Lys 9 histone methyltransferase (Suv39), which is enriched at heterochromatin and seems to be involved in gene repression and chromosome pairing. In addition, there are two mouse SET-domain-containing Suv39 histone methyltransferases, Suv39h1 and Suv39h2. Suv39h2 is preferentially expressed in the testis, and accumulates with chromatin of the sex body (Box 1). Indeed, mice mutated for both Suv39h1 and Suv39h2 are infertile owing to spermatogenic arrest⁶⁷. Finally, although the relationship between DNA methylation and DNA repair is not fully understood⁶⁸, HR23B-deficient animals show a phenotype that is similar to Suv39-null mice⁶⁹. HR23B is a ubiquitin-conjugating DNA repair enzyme that is involved in meiosis.

A fascinating possibility is that non-coding RNAs may have a function in spermatogenesis. For instance, naturally occurring RNAi constitute a powerful route to dynamically silence specific gene expression, so it is conceivable that such mechanisms may induce silencing initiation ahead of the more classical heterochromatinization process that is mediated by histone methyltransferase-mediated Lys 9 histone H3 methylation⁷⁰. It will be interesting to know how much specific RNAi is made in germ cells, and how the levels are controlled during differentiation and meiosis. Are there types of RNAi that can specifically target a whole chromosome? While these remain open questions, it is worth mentioning that germ cells contain an intriguing organelle, the chromatoid body (Box 1), which is mostly constituted by RNA. This cytoplasmic structure is conserved throughout evolution (the *Drosophila* equivalent is called nuage⁷¹). Its proteinaceous composition is still mostly unexplored. One component, however, is known and noteworthy: MVH (mouse homologue of VASA), an RNA helicase that is involved in RNA silencing⁷².

Histone modifications are recognized by bromodomain-containing proteins that specifically bind to acetylated lysines, and chromodomains that contain methylated Lys 9 on H3 ref. 73). One such protein is a testis-specific bromodomain protein called BRDT (bromodomain testis-specific), which induces chromatin remodelling in the presence of histone hyperacetylation⁷⁴. BRDT may mediate nuclear reorganization in spermiogenesis at the stage where histone hyperacetylation precedes the replacement of the testis-specific nuclear proteins⁷⁵.

What's coming?

Germ cell development is unique in the way that it generates the haploid cells that are responsible for the maintenance of the species. Understanding how epigenetic patterning occurs in the germ line may eventually help in the prevention of heritable diseases, improvement of assisted reproductive technologies, and stem cell therapy⁷⁶. The way is now paved for exciting new discoveries of the modifying enzymes that are responsible for these epigenetic modifications, the signalling pathways involved and the downstream effects on transcription, DNA repair and replication. Given that in males the spermatogenic process is continuous throughout life, it is

essential to elucidate how genetic and epigenetic processes are influenced by environmental cues. Finally, the long-standing link between metabolism and reproduction might also be based on epigenetic regulation, suggesting new research directions. New evidence suggests that embryonic DNA methylation patterns are influenced by maternal nutrition, and that epigenetic mutations induced by malnutrition could lead to development of diseases such as cancer and diabetes in adult life⁷⁷. □

doi:10.1038/nature03368.

1. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet.* **33** (suppl.), 245–254 (2003).
2. Cheung, P., Allis, C. D. & Sassone-Corsi, P. Signaling to chromatin through histone modifications. *Cell* **103**, 263–271 (2000).
3. Fischle, W., Wang, Y. & Allis, C. D. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**, 475–479 (2003).
4. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–453 (2003).
5. Sims, R. J. III, Nishioka, K. & Reinberg, D. Histone lysine methylation: a signature for chromatin function. *Trends Genet.* **19**, 629–639 (2003).
6. Bannister, A. J., Schneider, R. & Kouzarides, T. Histone methylation: dynamic or static? *Cell* **109**, 801–806 (2002).
7. Holliday, R. DNA methylation and epigenetic mechanisms. *Cell Biophys.* **15**, 15–20 (1989).
8. Sassone-Corsi, P. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science* **296**, 2176–2178 (2002).
9. DeBaun, M. R., Niemitz, E. L. & Feinberg, A. P. Association of *in vitro* fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19. *Am. J. Hum. Genet.* **72**, 156–160 (2003).
10. Rhind, S. M. *et al.* Human cloning: can it be made safe? *Nature Rev. Genet.* **4**, 855–864 (2003).
11. Sarma, K. & Reinberg, D. Histone variants meet their match. *Nature Rev. Mol. Cell Biol.* **6**, 1–11 (2005).
12. Marston, A. L. & Amon, A. Meiosis: cell-cycle controls shuffle and deal. *Nature Rev. Mol. Cell Biol.* **5**, 983–997 (2004).
13. Meistrich, M. L., Mohapatra, B., Shirley, C. R. & Zhao, M. Roles of transition nuclear proteins in spermiogenesis. *Chromosoma* **111**, 483–488 (2003).
14. Yu, Y. E. *et al.* Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. *Proc. Natl Acad. Sci. USA* **97**, 4683–4688 (2000).
15. Zhao, M. *et al.* Targeted disruption of the transition protein 2 gene affects sperm chromatin structure and reduces fertility in mice. *Mol. Cell Biol.* **21**, 7243–7255 (2001).
16. Meetei, A. R., Ullas, K. S., Vasupradha, V. & Rao, M. R. Involvement of protein kinase A in the phosphorylation of spermatid protein TP2 and its effect on DNA condensation. *Biochemistry* **41**, 185–195 (2002).
17. Sung, M. T. & Dixon, G. H. Modification of histones during spermiogenesis in trout: a molecular mechanism for altering histone binding to DNA. *Proc. Natl Acad. Sci. USA* **67**, 1616–1623 (1970).
18. Lahn, B. T. *et al.* Previously uncharacterized histone acetyltransferases implicated in mammalian spermatogenesis. *Proc. Natl Acad. Sci. USA* **99**, 8707–8712 (2002).
19. Oliva, R. & Dixon, G. H. Vertebrate protamine genes and the histone-to-protamine replacement reaction. *Progr. Nucleic Acid Res. Mol. Biol.* **40**, 25–94 (1991).
20. Lewis, J. D. *et al.* Histone H1 and the origin of protamines. *Proc. Natl Acad. Sci. USA* **101**, 4148–4152 (2004).
21. Pogany, G. C., Corzett, M., Weston, S. & Balhorn, R. DNA and protein content of mouse sperm. Implications regarding sperm chromatin structure. *Exp. Cell Res.* **136**, 127–136 (1981).
22. Cho, C. *et al.* Haploinsufficiency of protamine-1 or -2 causes infertility in mice. *Nature Genet.* **28**, 82–86 (2001).
23. Wu, J. Y. *et al.* Spermiogenesis and exchange of basic nuclear proteins are impaired in male germ cells lacking Camk4. *Nature Genet.* **25**, 448–452 (2000).
24. Xu, X., Toselli, P. A., Russell, L. D. & Seldin, D. C. Globozoospermia in mice lacking the casein kinase II α catalytic subunit. *Nature Genet.* **23**, 118–121 (1999).
25. Lee, K., Haugen, H. S., Clegg, C. H. & Braun, R. E. Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. *Proc. Natl Acad. Sci. USA* **92**, 12451–12455 (1995).
26. Grimes, S. *et al.* A rat histone H4 gene closely associated with the testis-specific H1t gene. *Exp. Cell Res.* **173**, 534–545 (1987).
27. Redon, C. *et al.* Histone H2A variants H2AX and H2AZ. *Curr. Opin. Genet. Dev.* **12**, 162–169 (2002).
28. Pandey, N. B. & Marzluff, W. F. The stem-loop structure at the 3' end of histone mRNA is necessary and sufficient for regulation of histone mRNA stability. *Mol. Cell Biol.* **7**, 4557–4559 (1987).
29. Moss, S. B., Challoner, P. B. & Groudine, M. Expression of a novel histone 2B during mouse spermiogenesis. *Dev. Biol.* **133**, 83–92 (1989).
30. Albig, W. *et al.* All known human H1 histone genes except the H1(0) gene are clustered on chromosome 6. *Genomics* **16**, 649–654 (1993).
31. Kobor, M. S. *et al.* Protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. *PLoS Biol.* **2**, E131 (2004).
32. Tagami, H., Ray-Gallet, D., Almouzni, G. & Nakatani, Y. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* **116**, 51–61 (2004).
33. Bramlage, B., Kosciessa, U. & Doenecke, D. Differential expression of the murine histone genes H3.3A and H3.3B. *Differentiation* **62**, 13–20 (1997).
34. Zalensky, A. O. *et al.* Human testis/sperm-specific histone H2B (hTSH2B). Molecular cloning and characterization. *J. Biol. Chem.* **277**, 43474–43480 (2002).
35. Martianov, I. *et al.* Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis. *Proc. Natl Acad. Sci. USA* **102**, 2808–2813 (2005).
36. Henikoff, S., Ahmad, K., Platano, J. S. & van Steensel, B. Heterochromatic deposition of centromeric histone H3-like proteins. *Proc. Natl Acad. Sci. USA* **97**, 716–721 (2000).

37. McKittrick, E., Gafken, P. R., Ahmad, K. & Henikoff, S. Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc. Natl Acad. Sci. USA* **101**, 1525–1530 (2004).
38. Crosio, C. *et al.* Mitotic phosphorylation of histone H3: spatio-temporal regulation by mammalian Aurora kinases. *Mol. Cell. Biol.* **22**, 874–885 (2002).
39. Zeitlin, S. G., Shelby, R. D. & Sullivan, K. F. CENP-A is phosphorylated by Aurora B kinase and plays an unexpected role in completion of cytokinesis. *J. Cell Biol.* **155**, 1147–1157 (2001).
40. Zalensky, A. O. *et al.* Human testis/sperm-specific histone H2B (hTSH2B). Molecular cloning and characterization. *J. Biol. Chem.* **277**, 43474–43480 (2002).
41. Niino, Y. S. *et al.* PKC δ II, a new isoform of protein kinase C specifically expressed in the seminiferous tubules of mouse testis. *J. Biol. Chem.* **276**, 36711–36717 (2001).
42. Tseng, T. C., Chen, S. H., Hsu, Y. P. & Tang, T. K. Protein kinase profile of sperm and eggs: cloning and characterization of two novel testis-specific protein kinases (AIE1, AIE2) related to yeast and fly chromosome segregation regulators. *DNA Cell Biol.* **17**, 823–833 (1998).
43. Park, E. J., Chan, D. W., Park, J. H., Oettinger, M. A. & Kwon, J. DNA-PK is activated by nucleosomes and phosphorylates H2AX within the nucleosomes in an acetylation-dependent manner. *Nucleic Acids Res.* **31**, 6819–6827 (2003).
44. Burma, S., Chen, B. P., Murphy, M., Kurimasa, A. & Chen, D. J. ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *J. Biol. Chem.* **276**, 42462–42467 (2001).
45. Shroff, R. *et al.* Distribution and dynamics of chromatin modification induced by a defined DNA double-strand break. *Curr. Biol.* **14**, 1703–1711 (2004).
46. Celeste, A. *et al.* Genomic instability in mice lacking histone H2AX. *Science* **296**, 922–927 (2002).
47. Barlow, C. *et al.* Atm deficiency results in severe meiotic disruption as early as leptotema of prophase I. *Development* **125**, 4007–4017 (1998).
48. Abraham, R. T. Checkpoint signaling: epigenetic events sound the DNA strand-breaks alarm to the ATM protein kinase. *Bioessays* **25**, 627–630 (2003).
49. McKee, B. D. & Handel, M. A. Sex chromosomes, recombination, and chromatin conformation. *Chromosoma* **102**, 71–80 (1993).
50. Chadwick, B. P. & Willard, H. F. Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. *Hum. Mol. Genet.* **10**, 1101–1113 (2001).
51. Baarends, W. M., Roest, H. P. & Grootegeed, J. A. The ubiquitin system in gametogenesis. *Mol. Cell. Endocrinol.* **151**, 5–16 (1999).
52. Hendzel, M. J., Lever, M. A., Crawford, E. & Th'ng, J. P. The C-terminal domain is the primary determinant of histone H1 binding to chromatin *in vivo*. *J. Biol. Chem.* **279**, 20028–20034 (2004).
53. Fantz, D. A. *et al.* Mice with a targeted disruption of the H1t gene are fertile and undergo normal changes in structural chromosomal proteins during spermiogenesis. *Biol. Reprod.* **64**, 425–431 (2001).
54. Doenecke, D. *et al.* Histone gene expression and chromatin structure during spermatogenesis. *Adv. Exp. Med. Biol.* **424**, 37–48 (1997).
55. Nayernia, K. *et al.* Male mice lacking three germ cell expressed genes are fertile. *Biol. Reprod.* **69**, 1973–1978 (2003).
56. Yan, W., Ma, L., Burns, K. H. & Matzuk, M. M. HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis. *Proc. Natl Acad. Sci. USA* **100**, 10546–10551 (2003).
57. Contreras, A. *et al.* The dynamic mobility of histone H1 is regulated by cyclin/CDK phosphorylation. *Mol. Cell. Biol.* **23**, 8626–8636 (2003).
58. Kuzmichev, A., Jenwein, T., Tempst, P. & Reinberg, D. Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol. Cell* **14**, 183–193 (2004).
59. Kim, J. M., Liu, H., Tazaki, M., Nagata, M. & Aoki, F. Changes in histone acetylation during mouse oocyte meiosis. *J. Cell Biol.* **7**, 37–46 (2003).
60. De La Fuente, R., Viveiros, M. M., Wigglesworth, K. & Eppig, J. J. ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes. *Dev. Biol.* **272**, 1–14 (2004).
61. Jedrusik, M. A. & Schulze, E. A single histone H1 isoform (H1.1) is essential for chromatin silencing and germline development in *Caenorhabditis elegans*. *Development* **128**, 1069–1080 (2001).
62. Clarke, H. J., Bustin, M. & Oblin, C. Chromatin modifications during oogenesis in the mouse: removal of somatic subtypes of histone H1 from oocyte chromatin occurs post-natally through a post-transcriptional mechanism. *J. Cell Sci.* **110**, 477–487 (1997).
63. Debey, P. *et al.* Competent mouse oocytes isolated from antral follicles exhibit different chromatin organization and follow different maturation dynamics. *Mol. Reprod. Dev.* **36**, 59–74 (1993).
64. Teranishi, T. *et al.* Rapid replacement of somatic linker histones with the oocyte-specific linker histone H1Foo in nuclear transfer. *Dev. Biol.* **266**, 76–86 (2004).
65. Buaa, F. W. *et al.* Plzf is required in adult male germ cells for stem cell self-renewal. *Nature Genet.* **36**, 647–652 (2004).
66. Costoya, J. A. *et al.* Essential role of Plzf in maintenance of spermatogonial stem cells. *Nature Genet.* **36**, 653–659 (2004).
67. Peters, A. H. *et al.* Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323–337 (2001).
68. Sedgwick, B. Repairing DNA-methylation damage. *Nature Rev. Mol. Cell Biol.* **5**, 148–157 (2004).
69. Roest, H. P. *et al.* Inactivation of the HR6B ubiquitin-conjugating DNA repair enzyme in mice causes male sterility associated with chromatin modification. *Cell* **86**, 799–810 (1996).
70. Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
71. Hay, B., Ackerman, L., Barbel, S., Jan, L. Y. & Jan, Y. N. Identification of a component of *Drosophila* polar granules. *Development* **103**, 625–640 (1988).
72. Toyooka, Y. *et al.* Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. *Mech. Dev.* **93**, 139–149 (2000).
73. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
74. Pivrot-Pajot, C. *et al.* Acetylation-dependent chromatin reorganization by BRDT, a testis-specific bromodomain-containing protein. *Mol. Cell. Biol.* **23**, 5354–5365 (2003).
75. Hazzouri, M. *et al.* Regulated hyperacetylation of core histones during mouse spermatogenesis: involvement of histone deacetylases. *Eur. J. Cell Biol.* **79**, 950–960 (2000).
76. Brinster, R. L. Germline stem cell transplantation and transgenesis. *Science* **296**, 2174–2176 (2002).
77. Waterland, R. A. & Jirtle, R. L. Early nutrition, epigenetic changes at transposons and imprinted genes, and enhanced susceptibility to adult chronic diseases. *Nutrition* **20**, 63–68 (2004).
78. Moss, S. B. & Orth, J. M. Localization of a spermatid-specific histone 2B protein in mouse spermiogenic cells. *Biol. Reprod.* **48**, 1047–1056 (1993).

Acknowledgements We were unfortunately unable to include all the relevant references owing to space constraints. We are grateful to C. D. Allis, D. Reinberg, R. Jaenisch, E. Borrelli, I. Davidson, N. Kotaja, U. Kolthur, G. Fienga, K. Hogeveen, C. Krausz, M. Parvinen, S. Henikoff, R. L. Brinster and all members of the Sassone-Corsi laboratory for critical reading of the manuscript, advice and stimulating discussions. S.K. is supported by fellowships from the Fondation pour la Recherche Médicale and the Marie Curie Programme.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.S.-C. (paolosc@igbmc.u-strasbg.fr).

An olivine-free mantle source of Hawaiian shield basalts

Alexander V. Sobolev^{1,2}, Albrecht W. Hofmann¹, Stephan V. Sobolev^{3,4} & Igor K. Nikogosian^{5,6}

¹Max-Planck-Institut für Chemie, Postfach 3060, 55020 Mainz, Germany

²Vernadsky Institute of Geochemistry, Russian Academy of Sciences, Kosygin street 19, 117975 Moscow, Russia

³GeoForschungsZentrum, Telegrafenberg E, D-14473, Potsdam, Germany

⁴Institute of Physics of the Earth, Russian Academy of Sciences, B. Gruzinskaya street 10, 123995 Moscow, Russia

⁵Faculty of Geosciences, Department of Petrology, Utrecht University, Budapestlaan 4, 3584 CD Utrecht, The Netherlands

⁶Faculty of Earth and Life Sciences Department of Petrology, Vrije Universiteit, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands

More than 50 per cent of the Earth's upper mantle consists of olivine and it is generally thought that mantle-derived melts are generated in equilibrium with this mineral. Here, however, we show that the unusually high nickel and silicon contents of most parental Hawaiian magmas are inconsistent with a deep olivine-bearing source, because this mineral together with pyroxene buffers both nickel and silicon at lower levels. This can be resolved if the olivine of the mantle peridotite is consumed by reaction with melts derived from recycled oceanic crust, to form a secondary pyroxenitic source. Our modelling shows that more than half of Hawaiian magmas formed during the past 1 Myr came from this source. In addition, we estimate that the proportion of recycled (oceanic) crust varies from 30 per cent near the plume centre to insignificant levels at the plume edge. These results are also consistent with volcano volumes, magma volume flux and seismological observations.

The upper mantle consists largely of peridotite with more than 50% of the mineral olivine¹. Consequently, nearly all petrological studies of mantle melting^{2–8} (with rare exceptions^{9,10}) start with the assumption that primary melt forms in equilibrium with olivine. This condition enforces strong constraints on the composition of the primary melt: at high pressure (more than 3.0 GPa), melts derived from garnet lherzolite are buffered at high MgO (more than 16%) and low SiO₂ (less than 47%). Higher SiO₂ contents (up to 48%) are possible only at higher degrees of melting, when garnet and high-Ca pyroxene are consumed and only olivine and low-Ca pyroxene persist^{6,8}. Nickel concentrations in the melt are constrained even more strongly, because olivine has the highest partition coefficient for Ni of any mantle silicate. This coefficient depends strongly on melt composition, generally decreasing with increasing olivine component¹¹ and sulphur content¹² in the melt. The latter is buffered at low levels by equilibrium with a sulphide phase in the source at high pressures¹³. Thus, liquids in equilibrium with typical mantle peridotite¹ can contain more than 800 p.p.m. Ni only if they are extremely enriched in olivine component (MgO > 22%), similarly to komatiites. For more realistic compositions of primary tholeiitic melts, with 16–18% MgO, such as those proposed for Hawaii⁴, the maximum Ni content will be about 500–600 p.p.m.

Hawaiian lavas represent the most productive active mantle plume. All volcanoes produced by this plume pass through three main stages: an alkalic pre-shield, tholeiitic shield-building, and an alkalic post-shield phase¹⁴. The shield phase produces about 95% of the total volcanic volume. Like most ocean island basalts, Hawaiian lavas are relatively enriched in incompatible elements (for example Ba, Nb, Ti and light rare-earth elements) and depleted in elements compatible with garnet (for example, heavy rare-earth elements). These characteristics are commonly attributed to low degrees of melting of fertile garnet lherzolite mantle requiring pressures of at least 2.5–3.0 GPa. However, in contrast with many other ocean island basalts, most Hawaiian lavas are relatively enriched in SiO₂, even at high MgO concentrations, and this combination is inconsistent with an equilibrium assemblage of olivine, pyroxenes and garnet^{2,7}. Three main hypotheses^{2,4,8} have been offered to resolve this dilemma. Here we show that the combination of high Ni and high Si

contents of parental Hawaiian magmas (from the main, shield-building stage) is incompatible with all these models, and this makes all models involving direct peridotite melting beneath Hawaii implausible. Instead, we propose that a model in which an olivine-free source component contributes 40–60% of the melts is consistent with geochemical and geophysical observations.

Ni in Hawaiian olivines and melts

To reconstruct the composition of the parental magma, we consider the earliest-formed, most magnesian olivines. Figure 1 shows relationships between Ni and forsterite (Fo) content in olivines from mantle rocks and several suites of high-Mg oceanic basalts. Assuming that olivine-melt partitioning for nickel and magnesium does not strongly vary with pressure⁵, compositions of source olivines and of the earliest olivines formed from the melts at shallow depths should be similar. Indeed, as expected, the Ni–Fo correlations are similar in high-Mg olivines from mid-ocean ridge basalts and from abyssal peridotites. Olivine suites from Iceland, Azores, Reunion, West Greenland, Gorgona komatiites and most Canaries olivines also fall in the same range. In contrast, olivines from shield-stage Hawaiian basalts show enormous ranges of NiO contents at a given Fo content, with a majority considerably enriched in NiO. Olivines from Hawaiian pre-shield and post-shield lavas are systematically lower in NiO (Fig. 1) and are consistent with common mantle peridotite sources. It is conceivable that there is a temperature effect in addition to, but difficult to separate from, the compositional effect on Ni partitioning. In that case, the high nickel content of Hawaiian olivines might simply be due to adiabatic cooling of melts initially generated at high temperature and pressure. If that were true, one would expect a similar effect for all equally high pressure–temperature melts such as Gorgona komatiites or west Greenland picrites⁸. However, Gorgona and west Greenland olivines show that this effect, if present at all, is too small to explain the Ni excess in Hawaiian olivines (Fig. 1).

The extreme Ni content in Hawaiian olivines (from shield-building lavas) suggests either that typical olivine-rich mantle lithologies are not sources of Hawaiian parental melts or that the melts have changed their composition after extraction.

The Ni content of crystallizing olivine depends mostly on the Ni

content and major element composition of the melt¹¹. We determine these characteristics from microscopic melt inclusions in olivine phenocrysts¹⁵. Figure 2 shows parental melts estimated from compositions of such inclusions by correction for early olivine crystallization (see Methods). The shield-stage parental melt compositions have very high Ni contents (700–1,000 p.p.m.) for a given MgO value, the most extreme being Koolau melts. In contrast, the single estimate for parental melt of an alkalic pre-shield suite (Loihi) has only 600 p.p.m. Ni.

Figure 2a also shows fields for lava compositions from the main stages. These are consistent with the estimated parental melt compositions, considering that the lavas have been variably fractionated by olivine loss and addition. However, the pre-shield and post-shield lavas have consistently lower Ni values for a given MgO content than the shield lavas. We conclude, from both the calculated parental melts and the high-Mg lava data, that the ‘Ni excess’ in Hawaiian magmas is a primary feature of the shield-stage tholeiites.

Finally, Fig. 2a shows the relationship of Ni against MgO in partial melts of fertile mantle peridotite in the pressure range 3.0–5.0 GPa, calculated from phase compositions and proportions in melting experiments⁶ and experimentally determined Ni partitioning between corresponding phases and melt (Supplementary Information). These results and similar estimates⁸ demonstrate the efficient buffering of Ni contents in all systems containing olivine and having a bulk Ni content of 1,900 p.p.m. (ref. 1). To shift this buffer to fit most Ni-rich Hawaiian shield magmas, the bulk Ni content of the peridotite would have to be more than 3,500 p.p.m., which is far in excess of Ni values observed in mantle lherzolite¹. Figure 2b shows analogous relationships for Ni/MgO ratios and SiO₂. Calculated parental melt compositions and lavas show an

overall positive correlation, with shield-stage magmas being consistently higher in SiO₂ than pre-shield and post-shield magmas. The solid lines representing the peridotite melting relationship are inconsistent with the calculated parental melts.

Problems with current Hawaiian plume models

Here we briefly review current petrological and geodynamic models explaining the compositional and geophysical features of Hawaiian volcanism. We show that none of these models is consistent with all the observations discussed above.

First, melts formed by high degrees of melting at high pressures in equilibrium with harzburgitic residues^{3,8}; this model is based on the assumption that the parental melts of Hawaiian magmas are highly magnesian (MgO ≥ 20%) and that high SiO₂ contents of the less magnesian lavas are produced by olivine fractionation. It explains moderately high Si in melts but requires unrealistically high Ni in the parental melt (for example, the trend given by the white circles in Fig. 2).

Second, high-pressure melts formed in equilibrium with garnet lherzolite are reequilibrated with lithospheric harzburgite at shallower depths^{2,7}; this model easily explains high Si contents and is consistent with both high Ti (and other incompatible elements) contents and with the ‘garnet signature’. However, it completely fails to explain high Ni contents of Hawaiian primary olivines and melts, because oceanic lithospheric olivines are much lower in Ni than olivines from Hawaiian shield tholeiites (Fig. 1).

Third, Hawaiian magmas are mixtures of low- and high-Si melts derived from lherzolite and eclogite, respectively^{4,9,16,17}; this model can explain the excess Si and is consistent with high concentrations of incompatible elements in the melts. However, it fails to reproduce

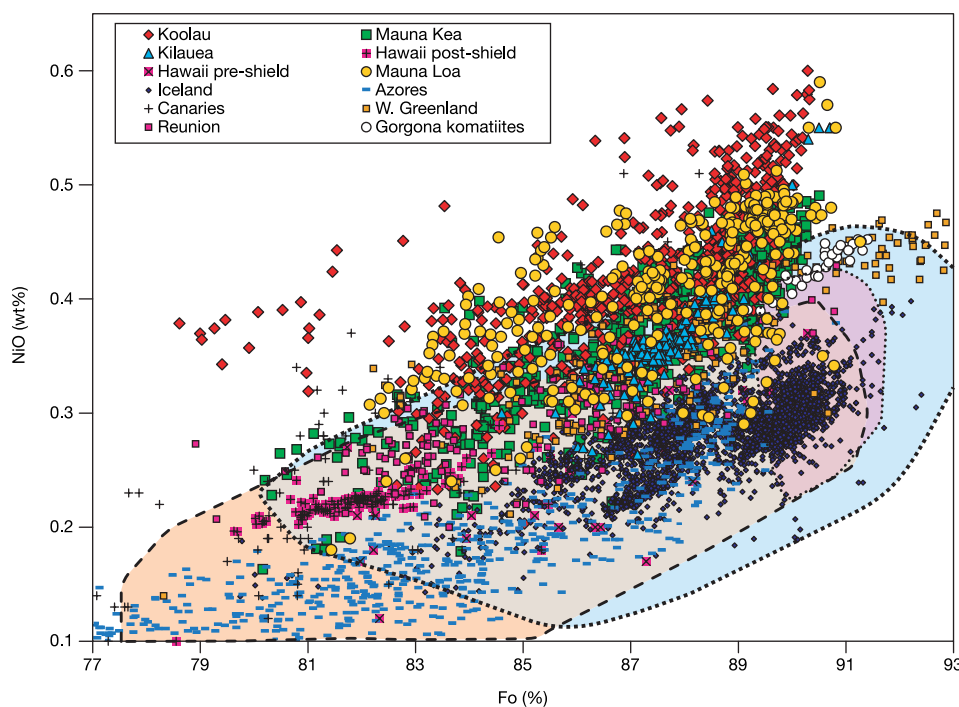


Figure 1 Compositions of olivines from mantle-derived rocks. Blue field, peridotites from mantle xenoliths, orogenic massifs and ophiolites; purple field, oceanic abyssal peridotites; beige field, phenocrysts from mid-ocean-ridge basalts; light green field, overlap between peridotite and phenocryst fields; pink field, overlap between oceanic abyssal peridotites and phenocrysts from mid-ocean-ridge basalts. Most data are from our unpublished database (data of A.V.S. on Hawaii, D. Kuzmin on Iceland, V. Kamenetsky on Gorgona, I. Nikogosian and T. Elliott on the Azores, I. Nikogosian on the Canaries and Reunion and V. Batanova for olivines from mantle peridotites). Olivines of Archæan

komatiites from Belingwe show NiO contents only 0.02 wt% higher than Gorgona komatiites (L. Danyushevsky, personal communication) and follow the upper boundary of the mantle peridotite field (blue). Additional data are from the GEOROC and PETDB databases⁴⁶ (see Supplementary Information for major references) and from ref. 47. Olivines from shield-stage Hawaiian basalts vary significantly in Ni content at constant Fo, with the majority systematically enriched in Ni compared with olivine from mantle peridotites, komatiites and common basalts. Olivines from post-shield and pre-shield Hawaiian basalts are similar to peridotites and common basalts.

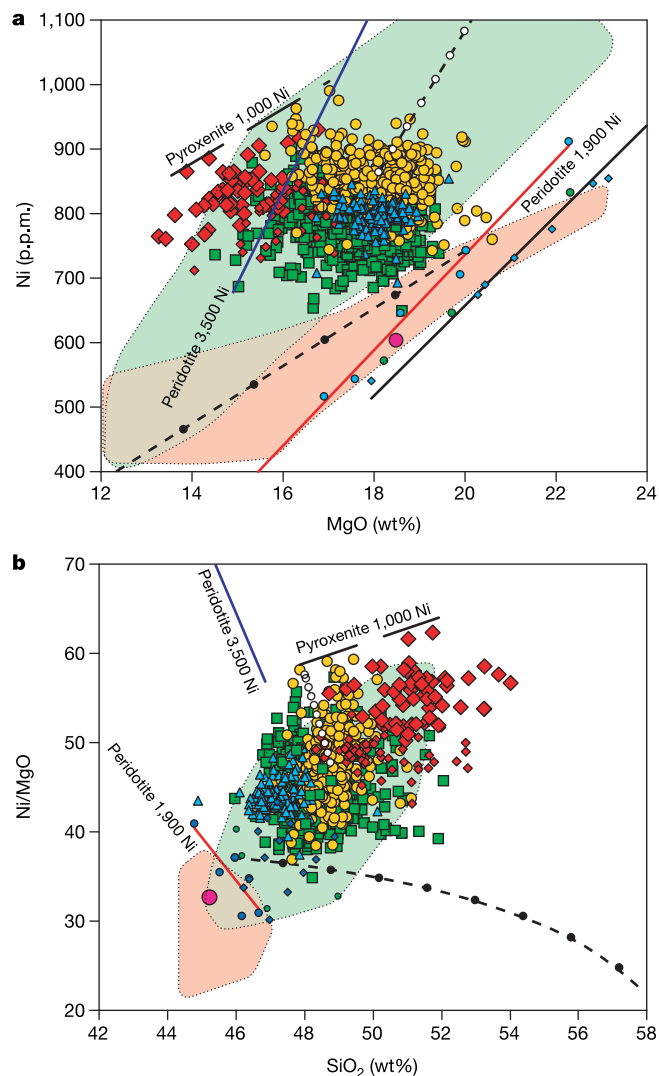


Figure 2 Parental melt and lava compositions, showing that Hawaiian shield parental melts are higher in Ni and Si than permitted in equilibrium with an olivine-bearing source, and that conventional models cannot explain this feature. Large symbols (Fig. 1) represent calculated compositions of parental Hawaiian melts. Large and small red diamonds correspond to parental melts for Makapuu and underlying series of Koolau Volcano, respectively. The large purple dot represents the parental magma composition of an alkaline glass from Loihi. The solid red line labelled 'peridotite 1,900 Ni' represents compositions of melts (small blue circles) in equilibrium with fertile mantle lherzolite in the pressure range 3.0–5.0 GPa, calculated for a bulk-mantle Ni content of 1,900 p.p.m. from melting experiments⁶ (Supplementary Table S1). The small green circles correspond to melt compositions calculated for residual harzburgite⁶. The small blue diamonds and black line represent compositions of melts in equilibrium with harzburgite (using the same bulk Ni)⁸. The blue solid line labelled 'peridotite 3,500 Ni' represents hypothetical melts for an unrealistically high bulk-lherzolite Ni content of 3,500 p.p.m. The thick dashed line labelled 'pyroxenite 1,000 Ni' shows compositions of estimated melts in equilibrium with pyroxenite with a bulk Ni content of 1,000 p.p.m. (Supplementary Table S2). Dashed line with white circles, olivine fractionation trend (1 wt% step) for a hypothetical parental melt leading to the average Mauna Loa parental melt composition; dashed line with black dots, mixing hyperbola between melt in equilibrium with mantle lherzolite (bulk Ni content 1,900 p.p.m.) at 4.5 GPa and 1,620 °C (ref. 6) and a hypothetical, eclogite-derived, high-Si melt⁴ with SiO₂ = 60 wt%; MgO = 4.5 wt% and Ni = 50 p.p.m. (10 wt% step). **a**, Plot of Ni against MgO for shield lavas (green field) and post-shield lavas (light brown). **b**, Plot of Ni (p.p.m.) to MgO (wt%) ratios against SiO₂ for shield lavas (green field) and post-shield lavas (light brown) having MgO between 15 and 19%, representing the range of parental melt compositions.

high Ni contents in the mixed melts because both partial melts of lherzolite and eclogite are lower in Ni (for example, the trend given by the black dots in Fig. 2).

Fourth, Koolau magmas produced by melting of recycled eclogite⁹: this model explains the high Si content in lavas and the enhanced productivity of the plume but fails to explain high Ni and Mg contents of Koolau olivines and melts.

Fifth, excess Ni comes from the Earth's core¹⁸: input from the core has also been suggested by studies of Os isotopes¹⁹ and high Fe/Mn ratios in lavas²⁰ and is conceivable because Ni is a major core component¹. However, the required Ni excess is more than 1,500 p.p.m. (Fig. 2a), which corresponds to a 3% contribution of core material with about 5% Ni (ref. 1). This is three times the amount suggested by Os isotopes and Fe/Mn ratios^{19,20}. In addition, any input from the core should markedly increase the platinum-group element abundances in the Hawaiian source and melts. This effect could be suppressed in the melts only by an exceptionally large amount of residual sulphides; however, this should also suppress the abundance of copper in the melt. But platinum-group elements are not specifically enriched, nor is copper specifically depleted, in Hawaiian lavas (GEOROC database, <http://georoc.mpch-mainz.gwdg.de/georoc/>). Finally, the Koolau melts with the most extreme Ni/Mg and Fe/Mn ratios (Fig. 2b and ref. 20) do not show Os isotopic evidence for core material¹⁹. All these observations argue against a core source of Ni (or Fe) excess.

Last, the three-dimensional dynamic model of Hawaiian plume emplacement and melting²¹ is apparently inconsistent with the new estimate of magma volume flux beneath Hawaii^{22,23}, which is at least double the flux used in the model²¹. The model can be adapted to the new data either by lowering the lithospheric thickness from 90 km to less than 70 km at the plume axis, or by increasing the potential temperature of the plume from 1,600 °C to more than 1,700 °C. This temperature seems to be far too high; the lithospheric thickness under the plume axis has recently been determined as 100–110 km (ref. 24).

Pyroxenitic source model

Because none of the existing models can explain the combination of high nickel and high SiO₂ in Hawaiian tholeiites, we propose a new model (Fig. 3) with the following three essential elements. First, the rising plume contains eclogite bodies that start melting at about 190–180 km depth. Second, this high SiO₂ initial melt infiltrates into and reacts with the adjacent peridotite, thereby eliminating olivine and producing a solid pyroxenite. Last, both pyroxenite and unreacted peridotite melt at depths between 140 and 100 km, ultimately producing hybrid magmas by mixing in conduits and crustal magma chambers.

The plume originally consists of (at least) two different lithologies: recycled oceanic crust and peridotite²⁵. The recycled component is a SiO₂-oversaturated eclogite derived from a mixture of primitive oceanic basalts, oceanic gabbros and sediments (Supplementary Table S3). Because the solidus temperature of this eclogite is much lower than that of peridotite, the eclogite starts melting at higher pressures²⁶. This produces a high-Si liquid that is highly reactive with olivine-bearing peridotite²⁷. Therefore, as melt infiltrates the peridotite it converts it to a solid, olivine-free pyroxenite. Under conditions of local equilibrium, the olivine replacement forms a sharp front, which advances into the peridotite²⁸. The proportion of melt required to convert all olivine in typical lherzolite is between 40 wt% and 60 wt% (Supplementary Information). According to the theory of metasomatism²⁸ and experimental studies²⁹, this process permits no intermediate lithologies.

Near-fractional melting of eclogite and reaction of this melt with peridotite continues until the eclogite is too refractory for further melting (Supplementary Information). At this stage there will potentially be three different lithologies: olivine-free pyroxenite,

original peridotite not affected by eclogite-derived melt, and eclogitic refractory restite not able to melt further. Hawaiian magmas therefore originate from the mixing of melts produced from two lithologies.

As the plume rises, the secondary pyroxenite starts melting with a much higher production rate than normal peridotite^{9,10,30,31}. The unreacted peridotite also starts melting, but producing lower melt fractions. The nickel content in melts derived from the olivine-free pyroxenite will be sharply increased, because olivine no longer controls the bulk partition coefficient (Fig. 2 and Supplementary Table S2).

The model predicts the mixing proportions of the source end-members for different Hawaiian volcanoes (Supplementary Information). Figure 4a and Table 1 show that the proportion of pyroxenite-derived melt is highest for Koolau (more than 80%) and Mauna Loa (about 60%) parental melts, and lowest for pre-shield melts (less than 10%). Kilauea and Mauna Kea lavas are intermediate. Recalculated to volcano volumes (Table 1), this yields a mean

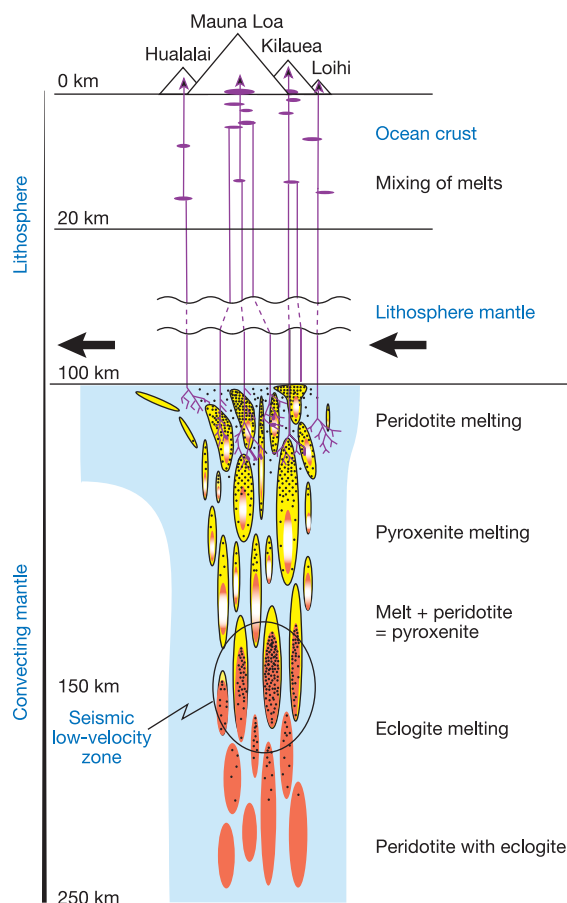


Figure 3 Model diagram of the Hawaiian mantle plume. Primary and secondary rock types are colour coded as follows: red, eclogite representing recycled oceanic crust; blue, peridotite; yellow, reaction (secondary) pyroxenite produced by infiltration of eclogite-derived melt into peridotite; white and red, eclogitic restite; black dots, melts; violet, magma pathways, conduits and small magma chambers. Recycled material is concentrated in the plume centre. The seismic low-velocity zone observed previously³⁶ in the depth range 170–130 km corresponds to significant melting of eclogite. This melt disappears at lower pressures because it separates from eclogite and is consumed by reaction with peridotite to produce secondary pyroxenite. Mixing of melts probably takes place at shallow crustal levels in small magma bodies rather than in the mantle or in large stable magma chambers. This is clearly seen from the very large variation in NiO and Fo contents of olivine within single Hawaiian picritic samples (Fig. 1) and melt inclusion data¹⁵.

contribution of about 50% of melt from the pyroxenitic source in the most recent (0.5–1.0-Myr-old) Hawaiian magma supply.

These mixing relationships can be tested by correlations between Os and Sr isotopes from Hawaiian shield lavas (Fig. 4b). This approach has previously been used to show mixing of melts rather

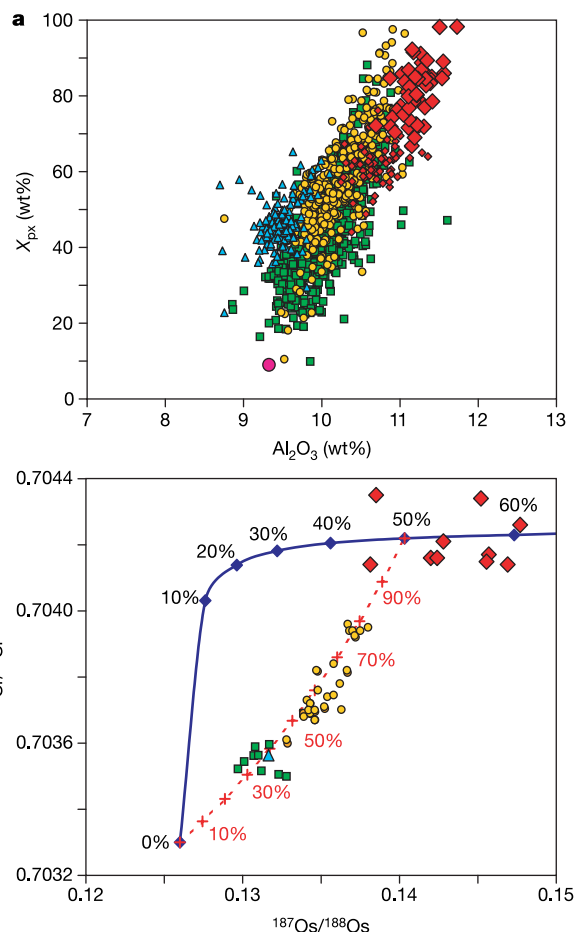


Figure 4 Hawaiian parental melts and lavas as mixtures of melts from two contrasting lithologies, olivine-free (reaction) pyroxenite and common peridotite. The symbols are the same as on Figs 1 and 2. **a**, Proportions of melt from pyroxenitic source (in wt%) plotted against Al_2O_3 for calculated parental melts. **b**, Mixing trajectories in Sr–Os isotope space. Data for Hawaiian shield lavas are from the GEOROC database, including refs 16, 32 and 48. Blue curve, mixing of eclogite-derived melt with peridotite; dashed red line, mixing of melts derived from peridotite and (reaction) pyroxenite, respectively. The parameters used for the blue line were as follows: mantle peridotite containing 3,000 p.p.t. Os with $^{187}Os/^{188}Os = 0.126$ and 10 p.p.m. Sr with $^{87}Sr/^{86}Sr = 0.7033$ was reacted with melt originating from eclogite having 50 p.p.t. Os with $^{187}Os/^{188}Os = 1.0$ and 300 p.p.m. Sr with $^{87}Sr/^{86}Sr = 0.70425$. The parameters used for the dashed red line were as follows: peridotite-derived melt having 400 p.p.t. Os and 300 p.p.m. Sr (the isotope ratios were the same as for unreacted peridotite); melt derived from pyroxenite (400 p.p.t. Os, 200 p.p.m. Sr, isotope ratios correspond to 50% peridotite reacted with 50% eclogite-derived melt, the proportions required to create an olivine-free lithology; see Supplementary Information). The observed isotopic compositions of Mauna Kea and Mauna Loa lavas correspond closely to the model of pure, binary melt mixing. This seems to rule out significant contributions from intermediate lithologies (for example, pyroxenite with residual, unreacted olivine), which should follow the strongly curved source-mixing hyperbola (blue line). The calculated proportions of the melt end-members are consistent in both **a** and **b** for Mauna Loa and Mauna Kea lavas, but Koolau data seem to be incompatible with the simple binary mixing model. This might be because of an additional (recycled) sedimentary component in Koolau^{4,49} and possibly because the initial amount of eclogite in the Koolau source was high enough to oversaturate the reaction zone with eclogite-derived melt.

than solid sources^{16,17,32}. The isotope data, with the exception of Koolau lavas, fit a mixing hyperbola between a melt derived from a pyroxenite created by the addition of 50% eclogite-derived melt (the amount required to react out all olivine) to the peridotite, and a melt resulting from the original peridotite. The relative proportions are similar to those shown in Fig. 4a.

We have modelled the above sequence of events by using incompatible element patterns for parental melts for several Hawaiian volcanoes. The results are shown in Supplementary Table S3 and Supplementary Fig. S1. In contrast with superficially similar earlier models¹⁵, the proportion of reactive melt from recycled crust (relative to ultramafic mantle) is very large (40–60%), to eliminate olivine. The melt fraction of the pyroxenitic source of Hawaiian lavas must therefore also be large (about 35% or greater), to match incompatible element contents and high Mg# ($\text{Mg\#} = 100\text{Mg}/(\text{Mg} + \text{Fe})$). In addition, the composition of the crustal component is somewhat more primitive than average mid-ocean-ridge basalt. The model successfully reproduces the trace element patterns of typical Mauna Loa parental melts and Sr-rich Mauna Loa melts¹⁵, as well as Kilauea and Loihi parental melts.

We now estimate the amount of recycled oceanic crust required to produce the final melts erupted in the different volcanoes (Supplementary Information). The results are shown in Table 1 and on a map of Hawaiian volcanoes (Fig. 5). The amounts of recycled crust range from 30% to nearly 10% in the central part of the plume and decrease to almost 0% at the plume edge. The results are clearly model dependent. Nevertheless, an independent, semiquantitative test is provided by the relative volumes of the volcanoes involved, which correlate with both the amount of pyroxenite melting and the amount of original eclogite (Table 1). These amounts are also consistent with the geophysical observations discussed below.

Geophysical consequences

The average content of eclogitic material in the bulk plume can be estimated as about 20% in the central part of the plume (radius 30 km) from the above estimates for specific volcanoes (Table 1 and

Fig. 5). A plume containing this amount of eclogite (with an assumed excess density of 2–3%)³³ will be buoyant in most of the mantle and will therefore be able to rise if its excess temperature is more than 200–300 °C.

The revised recent (younger than 2 Myr) Hawaiian magma flux is more than $10\text{ m}^3\text{ s}^{-1}$ (ref. 22) or close to $8\text{ m}^3\text{ s}^{-1}$ (ref. 23). In contrast, using our best estimates of potential temperature of 1,600 °C at the plume axis, an initial lithospheric thickness of 90 km and a maximum degree of peridotite melting of 15% (ref. 21), we obtain a magma flux of only $3.5\text{ m}^3\text{ s}^{-1}$ for a purely peridotitic plume. Note that in this and the following calculations (Supplementary Information), we estimate magma flux together with matching parameters of the Hawaii swell by using scaling from ref. 21. If we further consider the presence of up to 30% eclogite at the plume axis but ignore its effect on the density of the plume, we obtain a magma flux of 5–6 $\text{m}^3\text{ s}^{-1}$. This increased magma flux is caused by the addition of high-degree partial melt derived from pyroxenite. Finally, when we also consider that melting of eclogite at pressures of more than 4 GPa produces a garnet-rich restite³⁴, we obtain a total magma flux of more than $8.5\text{ m}^3\text{ s}^{-1}$, close to the observed flux^{22,23}. The reason for this higher estimate is that the eclogite restite significantly increases the bulk density of the plume (compared with a purely peridotitic plume), and hence a higher plume volume flux is required to support the swell²¹.

The presence of heavy eclogite restites completely compensates for the positive depletion buoyancy of the plume, which is significant in the models involving pure peridotite melting^{21,35}. The high swell topography in our model is therefore entirely supported by the higher temperature of the underlying plume material.

The model predicts strong partial melting of the eclogitic components in the central part of the plume. Highly viscous, Si-rich melts should remain in the residue until the degree of melting exceeds a threshold value (about 30%)²⁷ when it can infiltrate the surrounding peridotite. As melting proceeds during plume ascent, only the excess melt infiltrates the peridotite, but the threshold melt fraction gradually decreases because of decreasing SiO_2 content and

Table 1 Average compositions and formation parameters of estimated parental melts for recent Hawaiian volcanoes and Koolau

Parameter	Mauna Loa	Kilauea	Mauna Kea	Loihi alkaline	Average	PeM	PxM	Koolau Makapuu	Koolau KSDP
Age	0–80	0	460–550	0	<1,000	<1,000	<1,000	>2,000	2,600
V	80	30	33	0.66	143	69	74		
N	346	152	961	1	1,459	1,459	1,459	61	67
SiO ₂	48.8	47.3	48.3	45.2	48.4	47.6	49.1	51.3	50.2
TiO ₂	1.5	1.9	1.8	1.8	1.6	1.5	1.7	1.5	1.5
Al ₂ O ₃	10.2	9.5	10.1	9.3	10.0	9.0	10.9	11.3	10.8
Fe ₂ O ₃	1.2	1.2	1.2	1.4	1.2	1.2	1.2	1.1	1.1
FeO	9.6	10.0	9.8	10.3	9.8	9.9	9.6	9.0	9.3
MgO	18.1	18.1	17.6	18.5	18.0	20.8	15.3	15.0	16.8
CaO	8.0	9.2	8.6	10.1	8.4	7.4	9.2	7.0	7.3
Na ₂ O	1.7	1.6	1.6	1.7	1.6	1.5	1.8	2.4	2.0
K ₂ O	0.2	0.3	0.3	0.5	0.3	0.3	0.3	0.4	0.3
P ₂ O ₅	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.3
H ₂ O	0.3	0.4	0.3	0.6	0.3	0.3	0.3	0.5	0.3
Ni	858	797	779	603	827	716	928	836	824
For _{ol}	90.7	90.4	90.3	90.3	90.6	91.7	89.5	89.8	90.5
NiO _{ol}	0.55	0.51	0.52	0.37	0.53	0.32	0.76	0.72	0.60
T	1,369	1,364	1,357	1,367	1,363	1,413	1,315	1,314	1,347
X _{px}	0.57	0.46	0.45	0.09	0.52	0.00	1.00	0.81	0.62
X _{crc}	0.26	0.12	0.12	0.02	0.20			High	High
C _{pe}	0.61	0.82	0.83	0.98	0.71				
C _{er}	0.13	0.06	0.06	0.01	0.10				
F _{px}	0.45	0.35	0.35	0.25	0.41		0.41		
F _{pe}	0.15	0.06	0.06	0.04	0.11	0.11			

Ages are in kyr. Major elements (in wt%) and Ni (in p.p.m.) were calculated from melt inclusion and host olivine compositions, corrected for olivine fractionation (see Methods). V, volcano volume in $1,000\text{ km}^3$ (after (<http://hvo.wr.usgs.gov>)), volume of end-members calculated from their proportions (see below). N, number of melt inclusions (or glass for Loihi). For_{ol} and NiO_{ol}, forsterite (mol%) and NiO (wt%) contents of olivines in equilibrium with parental melts at 0.1 MPa and at the temperature indicated (T, in °C). The following parameters (all in weight fractions) are defined in Supplementary Information: degree of eclogite melting for all models ($F_e = 0.50$); proportion of eclogite-derived melt reacted with peridotite to form olivine-free pyroxenite for all models ($X_e = 0.50$); X_{px} , proportion of melt from pyroxenitic source; X_{crc} , amount of recycled crust (= eclogite); C_{pe} , weight fraction of unreacted peridotite; C_{er} , weight fraction of eclogitic restite; F_{px} , degree of melting of pyroxenite; F_{pe} , degree of melting of peridotite. Average, average parental melt for Mauna Loa, Mauna Kea and Kilauea, weighted by volcano volume. PeM and PxM, end-member compositions of peridotite-derived and pyroxenite-derived melts, respectively, each representing a volume-weighted average of end-member estimates for Mauna Loa, Mauna Kea and Kilauea (Supplementary Information). Peridotite-derived end-member is a high-Mg picrite similar to published estimates⁸ and is in equilibrium with peridotite under 100-km-thick Hawaiian lithosphere⁶. Pyroxenite-derived end-member has significantly higher Si, Al, Ca and Ni, and lower Mg contents, and is olivine undersaturated under Hawaiian lithosphere. There is a clear positive correlation between estimated amounts of recycled material (or proportion of pyroxenite-derived melt) and volcano volume. Parental melts from the Makapuu series of Koolau possess the highest Ni and Si contents and require an almost purely pyroxenitic source. Melts from the Koolau Scientific Drilling Project have compositions between those of Makapuu and Mauna Loa, supporting the conclusions of ref. 38.

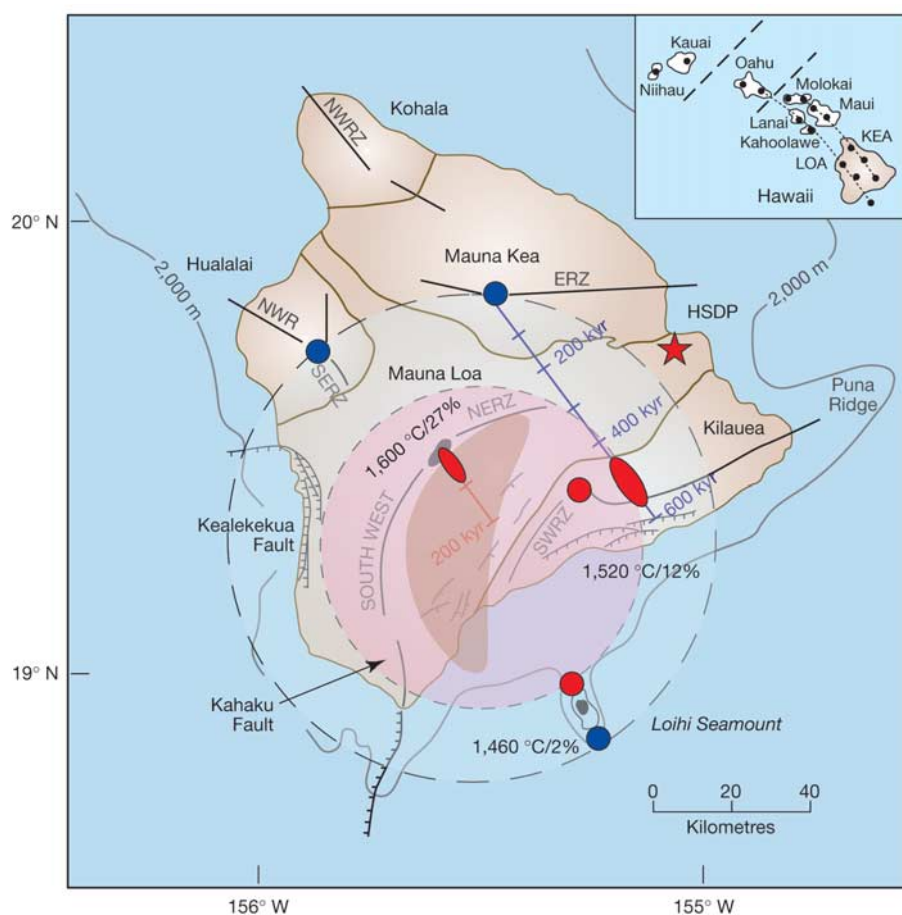


Figure 5 Schematic map of present position of the Hawaiian mantle plume, modified from ref. 50. Red and blue points and ellipses indicate sampled areas of plume corrected for eruption ages. Numbers near points show estimated potential plume temperatures and amounts (wt%) of recycled oceanic crust (see also Table 1, Methods and Supplementary Information). The pink circle shows the assumed position of the central zone of the plume. The light blue area represents the outer (and colder) portion of the plume. The dark

shadowed area (half-moon-shaped) indicates the position of the seismic low-velocity zone at depths between 170 and 130 km (ref. 36). Maximum amounts of recycled material and maximum temperatures are found under Mauna Loa and close to the plume centre. The average amount of recycled crust in the central (pink) zone is estimated to be about 20% (Table 1), whereas in the outer (blue) zone of the plume it decreases to 0%. The low-velocity zone coincides with the region of maximum amount of recycled crust.

melt viscosity. The high initial threshold should create a prominent seismic low-velocity zone at a depth range between the early stage (10% of partial melting) and the end of eclogite melting (Supplementary Information). These depths depend on the potential temperature of the plume and correspond to 170–130 km at a potential temperature of 1,600 °C. The seismic low-velocity zone has indeed been detected at this depth range below the southern part of the Big Island by using P-to-S converted waves³⁶.

Discussion

The proposed model explains several Hawaiian magma characteristics that were previously difficult to reconcile.

First, high and variable Ni contents are now consistent with high Si and with high incompatible element contents of MgO-rich parental melts, because olivine no longer buffers Ni and Si, and incompatible elements come from the recycled component and from low-fraction melting of peridotite.

Second, the high magma productivity of the plume is explained by the high proportion of pyroxenite, which produces much higher melt fractions than peridotite^{10,30,31}, and by an increased plume volume flux, required to support the Hawaiian swell if the plume contains heavy eclogite restite.

Third, the nearly linear correlation between Os and Sr isotopes is explained by the mixing of melts from only two stable lithologies, olivine-free reaction pyroxenite and unreacted peridotite, instead of melting variously mixed sources.

Fourth, the model explains the prominent seismic low-velocity zone in the depth range 170–130 km below the southern part of the Big Island³⁶, because a substantial amount of melt is retained in the eclogitic bodies in the central part of the plume.

Last, the model explains high Fe/Mn ratios of Hawaiian magmas²⁰, without an Fe contribution from the core, because in the absence of olivine the Fe/Mn ratio of the melt is expected to be higher than in the source²⁰.

A seemingly paradoxical aspect of the model is that Hawaiian tholeiites are universally rich in olivine, whereas we propose a very significant role of an olivine-free source. However, most parental melts are undersaturated with respect to olivine at depths greater than 60 km (refs 2, 7). Olivine saturation starts at lower pressures, and abundant olivine phenocrysts crystallize predominantly in crustal magma chambers^{3,15}.

One possibly troubling aspect is the fact that the Ni-rich, olivine-undersaturated parental melts must traverse the oceanic lithosphere and sublithospheric mantle without interacting with ambient olivine. This is therefore not consistent with currently popular models for melt transport through dunite channels in mid-ocean environments³⁷. Instead, we suggest that Hawaiian parental (primary) melts create pyroxenitic reaction channels for sublithospheric melt transport by a mechanism analogous to the formation of dunite channels³⁷, but remove olivine (rather than pyroxenes) from the peridotite. In the brittle lithosphere, the melts are likely to move through fractures so that the melt does not

interact with the wall rock. A low interaction between Hawaiian melts and wall rocks during transport has also been suggested on the basis of Os isotopes³² and compositions of melt inclusions¹⁵. The large variation in Ni contents in Hawaiian olivines and melts (Figs 1 and 2) suggests that different parental and primary melts mix in small crustal magma chambers and/or conduits.

Koolau exhibits an anomalous behaviour among Hawaiian volcanoes in having a maximum amount of pyroxenitic component at only moderate volcano volume. In addition, the amount of pyroxenitic component tends to increase towards the end of the shield stage of this volcano (Makapuu stage³⁸). As in ref. 9 we can explain this feature by the local input of an exceptionally large block of recycled eclogite. In this model, the moderate size of this volcano manifests a local 'shortage' of a peridotitic component. Inputs from peridotite melting will further decrease as the volcano moves to a colder part of the plume at the end of the shield stage.

The question arises whether the model described here also applies to other mantle plumes. Because a secondary pyroxenitic source derived from recycled oceanic crust will start to melt at a higher pressure than will peridotite^{9,10,30,31}, its effect (in the form of a high Ni–Si component) should be most clearly recognizable in plume basalts emplaced on thick lithosphere. Indeed, although the highest Ni excess in olivine occurs on Hawaii, relatively high values are also found on the Canaries and west Greenland (Fig. 1) and in flood basalts from Siberia¹⁸ and Karroo-Etendeka³⁹ (all with thick lithosphere), whereas low Ni values are found on Iceland and the Azores (thin lithosphere). □

Methods

Samples

We present a summary of data for host olivines and melt inclusions from the following Hawaiian shield picrites: (1) Mauna Loa, historic, 1868 (H-OC), prehistoric (H-1) and 50 kyr ago (R129–8.1) described in refs 3 and 15, and 80 kyr ago (SR118–8–8) from HSDP-2 site⁴⁰; (2) Mauna Kea, 450–550 kyr old (A.V.S., unpublished data on samples SR515–3.6, SR542–9.6, SR554–2.3, SR759–4.0, SR863–20.7, SR889–15.0, SR912–18.0, SR930–9.1, SR935–21.4, SR954–10.7, SR962–18.1 from HSDP-2 site⁴⁰); (3) Kilauea, Kilauea-Iki (A.V.S., unpublished data for samples IKI-22, IKI-44, K97–12 provided by A. T. Anderson); (4) Koolau (A.V.S., unpublished data from samples KooS10 from Makapuu series (provided by A. Rocholl) and R8–1.1–1197.1 core from Koolau Scientific Drilling Project (KSDP) site³⁸).

Olivine hosts and melt inclusions

Host olivine phenocrysts and melt inclusions were analysed for major elements with a Jeol JXA8200 electron probe at the Max Planck Institute for Chemistry, with standards⁴¹ and an extended counting time. Ni in olivines was analysed at 20 kV accelerating voltage and 20 nA current with a 120 s counting time, using NiO as a standard and San Carlos olivine (USNM 111312/444; ref. 41) for continuous monitoring. Typical external precisions are better than 1% relative (1 s.e.m.) for Ni in olivines, 0.1% relative for Fo in olivines and 0.5–1.0% relative for major elements in the melt inclusions.

Inclusions in olivines from all volcanoes are naturally quenched except samples H-OC from Mauna Loa and KooS10 from Koolau volcano. The latter were heated and quenched from a temperature of 1,250 °C at Vrije Universiteit, The Netherlands, and Vernadsky Institute, Russia, respectively.

All inclusion compositions were recalculated to equilibrium with host olivines, taking into account olivine crystallization on the walls and Fe–Mg redistribution with host mineral as described in ref. 15. The original FeO contents for the trapped melts were defined as a function of SiO₂ contents in the melt: $\text{FeO}_{\text{total}} = 24.93 - 0.2815 \times \text{SiO}_2$, based on the strong correlation of Hawaiian tholeiitic lavas with MgO contents between 10 wt% and 20 wt% (GEOROC database).

Ni in melts

Ni contents in olivine-hosted inclusions are severely lowered by olivine crystallization on the walls of cavities. Therefore, rather than measuring Ni in the melt inclusions we calculate Ni concentrations from measured Ni contents in the host olivine and distribution coefficients estimated for known compositions of melt and olivine with the use of the formulation in ref. 11. This formulation does not take into account the S content in the melt. However, Li *et al.*¹² showed that sulphur concentrations greater than 1,000 p.p.m. cause Ni–S complexing in the melt and thus produce excess Ni in the melt over that predicted by Beattie's model. Nevertheless, this average Ni excess is less than 70 p.p.m. for primitive glasses with S concentrations less than 1,100 p.p.m. Because S concentrations in most melts trapped in both high-magnesium Hawaiian olivines and primitive Hawaiian glasses are less than 1,000 p.p.m., and sulphide globules are normally absent as inclusions in high-Mg olivines⁴², no correction has been applied to the Ni distribution. Low S concentrations in most primitive mantle-derived melts, and their

apparent sulphur undersaturation at low pressures, can be well understood in view of the strong negative correlation of S saturation level with pressure¹³.

Reconstruction of parental melt compositions

The compositions of parental melts have been calculated from the compositions of corrected trapped melts or from magnesium glass composition for Loihi⁴³ by the back-calculation of olivine fractionation up to equilibrium with the highest-Mg olivine known for a given volcano. We assume that Fo contents of such olivine are defined as a function of its NiO content: $\text{Fo} = 93.12 - 5.35 \times \text{NiO}$ for Mauna Kea and Kilauea, and $\text{Fo} = 93.68 - 5.35 \times \text{NiO}$ for Mauna Loa and Koolau. These functions constrain the high-Fo corner of the Fo–NiO olivine diagram (Fig. 1) and are consistent with the fact that the pyroxenitic end-member should be highest in Ni and lowest in Mg#. Calculations were performed for inclusions in olivine (Fo > 86) with a model of Fe–Mg olivine–melt partitioning⁸ and an $\text{Fe}^{3+}/(\text{Fe}^{2+} + \text{Fe}^{3+})$ ratio of the melt equal to 0.10 (ref. 3). H₂O concentrations in parental melts were estimated by using data from ref. 42. Ni concentrations were calculated with the olivine–melt distribution formulation¹¹. Calculations were performed with PETROLOG software⁴⁴.

Estimation of potential mantle temperature

Available compositions of parental melts were used to constrain the potential mantle temperature of the Hawaiian plume, on the assumption that reconstructed melts separated from their sources at about 100 km, where the transition from lithosphere to asthenosphere is imaged under the Big Island²⁴ (with a 10-km-thick plume freezing zone²¹ this corresponds to the initial lithospheric thickness of 90 km). To constrain liquidus temperatures we used olivine–melt Mg–Fe partitioning⁸, a dT/dP slope of 45 °C/GPa, and compositions of parental melts with a minimum amount (less than 10%) of pyroxenite-derived melt. For 3.0–3.3 GPa pressure this yields, respectively, 1,550–1,565 °C for Mauna Loa, 1,520–1,535 °C for Mauna Kea and Kilauea, and 1,500–1,515 °C for alkaline Loihi. With parameterization from ref. 45, these temperatures correspond to ranges in potential mantle temperature of 1,620–1,570 °C for Mauna Loa, 1,520 °C for Mauna Kea and Kilauea, and 1,460 °C for Loihi. This yields a potential mantle temperature at the centre of the plume close to 1,600 °C, nearly identical to the estimates of ref. 21.

Received 15 July 2004; accepted 31 January 2005; doi:10.1038/nature03411.

- McDonough, W. F. & Sun, S. S. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
- Eggs, S. M. Petrogenesis of Hawaiian tholeiites. I. Phase-equilibria constraints. *Contrib. Mineral. Petrol.* **110**, 387–397 (1992).
- Sobolev, A. V. & Nikogosian, I. K. Petrology of long-lived mantle plume magmatism: Hawaii (Pacific) and Reunion Island (Indian Ocean). *Petrology* **2**, 111–144 (1994).
- Hauri, E. H. Major element variability in the Hawaiian mantle plume. *Nature* **382**, 415–419 (1996).
- Herzberg, C. & Zhang, J. Z. Melting experiments on anhydrous peridotite KLB. 1. Compositions of magmas in the upper mantle and transition zone. *J. Geophys. Res. Solid Earth* **101**, 8271–8295 (1996).
- Walter, M. J. Melting of garnet peridotite and the origin of komatiite and depleted lithosphere. *J. Petrol.* **39**, 29–60 (1998).
- Wagner, T. P. & Grove, T. L. Melt/harzburgite reaction in the petrogenesis of tholeiitic magma from Kilauea volcano, Hawaii. *Contrib. Mineral. Petrol.* **131**, 1–12 (1998).
- Herzberg, C. & O'Hara, M. J. Plume-associated ultramafic magmas of Phanerozoic age. *J. Petrol.* **43**, 1857–1883 (2002).
- Takahashi, E. & Nakajima, K. In *Hawaiian Volcanoes: Deep Underwater Perspectives* (eds Takahashi, E., Lipman, P. W., Garcia, O. M., Naka, J. & Aramaki, S.) 403–418 (Geophys. Monogr. 128, AGU, Washington DC, 2002).
- Hirschmann, M. M., Kogiso, T., Baker, M. B. & Stolper, E. M. Alkaline magmas generated by partial melting of garnet pyroxenite. *Geology* **31**, 481–484 (2003).
- Beattie, P., Ford, C. & Russell, D. Partition coefficients for olivine–melt and ortho-pyroxene–melt systems. *Contrib. Mineral. Petrol.* **109**, 212–224 (1991).
- Li, C., Ripley, E. M. & Mathez, E. A. The effect of S on the partitioning of Ni between olivine and silicate melt in MORB. *Chem. Geol.* **201**, 295–306 (2003).
- Mavrogenes, J. A. & O'Neill, H. S. C. The relative effects of pressure, temperature and oxygen fugacity on the solubility of sulfide in mafic magmas. *Geochim. Cosmochim. Acta* **63**, 1173–1180 (1999).
- Moore, J. G. & Clague, D. Volcano growth and evolution of island of Hawaii. *Geol. Soc. Am. Bull.* **104**, 1471–1484 (1992).
- Sobolev, A. V., Hofmann, A. W. & Nikogosian, I. K. Recycled oceanic crust observed in 'ghost plagioclase' within the source of Mauna Loa lavas. *Nature* **404**, 986–990 (2000).
- Lassiter, J. C. & Hauri, E. H. Osmium-isotope variations in Hawaiian lavas: evidence for recycled oceanic lithosphere in the Hawaiian plume. *Earth Planet. Sci. Lett.* **164**, 483–496 (1998).
- Kogiso, T., Hirschmann, M. M. & Reiners, P. W. Length scales of mantle heterogeneities and their relationship to ocean island basalt geochemistry. *Geochim. Cosmochim. Acta* **68**, 345–360 (2004).
- Ryabchikov, I. D. High NiO content in mantle-derived magmas as evidence for material transfer from the Earth's core. *Dokl. Earth Sci.* **389**, 437–439 (2003).
- Brandon, A. D., Norman, M. D., Walker, R. J. & Morgan, J. W. Os-186–Os-187 systematics of Hawaiian picrites. *Earth Planet. Sci. Lett.* **174**, 25–42 (1999).
- Humayun, M., Qin, L. P. & Norman, M. D. Geochemical evidence for excess iron in the mantle beneath Hawaii. *Science* **306**, 91–94 (2004).
- Ribe, N. M. & Christensen, U. R. The dynamical origin of Hawaiian volcanism. *Earth Planet. Sci. Lett.* **171**, 517–531 (1999).
- Vidal, V. & Bonneville, A. Variations of the Hawaiian hot spot activity revealed by variations in the magma production rate. *J. Geophys. Res. Solid Earth* **109**, B03104 (2004).
- Van Ark, E. & Lin, J. Time variation in igneous volume flux of the Hawaii–Emperor hot spot seamount chain. *J. Geophys. Res. Solid Earth* **109**, B11401 (2004).
- Li, X. Q., Kind, R., Yuan, X. H., Wolbern, I. & Hanka, W. Rejuvenation of the lithosphere by the Hawaiian plume. *Nature* **427**, 827–829 (2004).
- Hofmann, A. W. & White, W. M. Mantle plumes from ancient oceanic crust. *Earth Planet. Sci. Lett.* **57**, 421–436 (1982).
- Yasuda, A., Fujii, T. & Kurita, K. Melting phase relations of an anhydrous mid-ocean ridge basalt from

- 3 to 20 GPa: Implications for the behavior of subducted oceanic crust in the mantle. *J. Geophys. Res. Solid Earth* **99**, 9401–9414 (1994).
27. Yaxley, G. M. & Green, D. H. Reactions between eclogite and peridotite: mantle refertilisation by subduction of oceanic crust. *Schweiz. Mineral. Petrogr. Mitt.* **78**, 243–255 (1998).
28. Korzhinskii, D. S. *Theory of Metasomatic Zoning* (Clarendon Press, Oxford, 1970).
29. Yaxley, G. M., Sobolev, A. V. & Snow, J. High-pressure partial melting of gabbro and the preservation of 'ghost plagioclase' signatures. *Geochim. Cosmochim. Acta* **68**, A578 (2004).
30. Yaxley, G. M. Experimental study of the phase and melting relations of homogeneous basalt plus peridotite mixtures and implications for the petrogenesis of flood basalts. *Contrib. Mineral. Petrol.* **139**, 326–338 (2000).
31. Pertermann, M. & Hirschmann, M. M. Partial melting experiments on a MORB-like pyroxenite between 2 and 3 GPa: Constraints on the presence of pyroxenite in basalt source regions from solidus location and melting rate. *J. Geophys. Res. Solid Earth* **108**(B2), 2125 (2003).
32. Hauri, E. H. & Kurz, M. D. Melt migration and mantle chromatography. 2: a time-series Os isotope study of Mauna Loa Volcano, Hawaii. *Earth Planet. Sci. Lett.* **153**, 21–36 (1997).
33. Ono, S., Ito, E. & Katsura, T. Mineralogy of subducted basaltic crust (MORB) from 25 to 37 GPa, and chemical heterogeneity of the lower mantle. *Earth Planet. Sci. Lett.* **190**, 57–63 (2001).
34. Kogiso, T., Hirschmann, M. M. & Frost, D. J. High-pressure partial melting of garnet pyroxenite: possible mafic lithologies in the source of ocean island basalts. *Earth Planet. Sci. Lett.* **216**, 603–617 (2003).
35. Morgan, J. P., Morgan, W. J. & Price, E. Hotspot melting generates both hotspot volcanism and a hotspot swell. *J. Geophys. Res. Solid Earth* **100**, 8045–8062 (1995).
36. Li, X. *et al.* Mapping the Hawaiian plume conduit with converted seismic waves. *Nature* **405**, 938–941 (2000).
37. Kelemen, P. B., Shimizu, N. & Salters, V. J. M. Extraction of Mid-Ocean-Ridge Basalt from the Upwelling Mantle by Focused Flow of Melt in Dunite Channels. *Nature* **375**, 747–753 (1995).
38. Haskins, E. H. & Garcia, M. O. Scientific drilling reveals geochemical heterogeneity within the Ko'olau shield, Hawai'i. *Contrib. Mineral. Petrol.* **147**, 162–188 (2004).
39. Gibson, S. A., Thompson, R. N. & Dickin, A. P. Ferropicrites: geochemical evidence for Fe-rich streaks in upwelling mantle plumes. *Earth Planet. Sci. Lett.* **174**, 355–374 (2000).
40. Stolper, E. M., DePaolo, D. J. & Thomas, D. M. Introduction to special section: Hawaii Scientific Drilling Project. *J. Geophys. Res. Solid Earth* **101**, 11593–11598 (1996).
41. Yarosewich, E. J., Nelen, J. A. & Norberg, J. A. Reference sample for electron microprobe analysis. *Geostand. Newsl.* **4**, 43–47 (1980).
42. Hauri, E. SIMS analysis of volatiles in silicate glasses. 2: isotopes and abundances in Hawaiian melt inclusions. *Chem. Geol.* **183**, 115–141 (2002).
43. Garcia, M. O., Foss, D. J. P., West, H. B. & Mahoney, J. J. Geochemical and isotopic evolution of Loihi Volcano, Hawaii. *J. Petrol.* **36**, 1647–1674 (1995).
44. Danyushevsky, L. V. The effect of small amounts of H₂O crystallisation of mid-ocean ridge and backarc basin magmas. *J. Volcanol. Geotherm. Res.* **110**, 265–280 (2001).
45. Manglik, A. & Christensen, U. R. Effect of mantle depletion buoyancy on plume flow and melting beneath a stationary plate. *J. Geophys. Res. Solid Earth* **102**, 5019–5028 (1997).
46. Lehnert, K. A. *et al.* A global geochemical database: Application to mid-ocean ridges and ocean islands. *Eos* **79** (Fall Meet. Suppl.), F44 (1998).
47. Garcia, M. O. in *Hawaiian Volcanoes: Deep Underwater Perspectives* (eds Takahashi, E., Lipman, P. W., Garcia, O. M., Naka, J. & Aramaki, S.) 391–402 (Geophys. Monogr. 128, AGU, Washington DC, 2002).
48. Bennett, V. C., Esat, T. M. & Norman, M. D. Two mantle-plume components in Hawaiian picrites inferred from correlated Os-Pb isotopes. *Nature* **381**, 221–224 (1996).
49. Blichert-Toft, I., Frey, F. A. & Albarede, F. Hf isotope evidence for pelagic sediments in the source of Hawaiian basalts. *Science* **285**, 879–882 (1999).
50. Kurz, M. D., Curtice, J., Lott, D. E. & Solow, A. Rapid helium isotopic variability in Mauna Kea shield lavas from the Hawaiian Scientific Drilling Project. *Geochem. Geophys. Geosyst.* **5**, Q04G14 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the HSDP and KSDP teams, A.T. Anderson and A. Rocholl for providing samples; E. Yarosewich for supplying microprobe standards; B. Schulz-Dobrick for supervising the purchase and installation of the Jeol Jxa 8200 Electron Microprobe in the Max Planck Institute for Chemistry; E. Macsenaere-Riester for assistance with this; A. Gurenko and N. Groschopf for maintaining the microprobe; A. Yasevich, V. Sobolev and M. Kamenetsky for help in sample preparation; D. Kuzmin for help in electron probe analyses; D. Kuzmin, V. Kamenetsky and V. Batanova for providing unpublished olivine analyses; L. Danyushevsky for making access available to PETROLOG thermodynamic modelling software and for updating it for use with nickel; C. Herzberg, V. Kamenetsky, A. Gurenko, H. Dick, G. Woerner, C. Langmuir, P. Kelemen, L. Kogarko, I. Ryabchikov, A. Ariskin and D. DePaolo for discussions; and M. Garcia, L. Danyushevsky, S. Huang, M. Portnyagin, K. Putirka and D. Canil for comments that improved the clarity of the manuscript. This work was supported by a Wolfgang Paul Award of the Alexander von Humboldt Foundation to A.V.S. Partial support was received from the Russian Academy of Science and Russian Federation President's grants to A.V.S. and from ISES (Netherlands Research Centre for Integrated Solid Earth Science) to I.K.N.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.V.S. (asobolev@mpch-mainz.mpg.de).

CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair

Fumiko Esashi¹, Nicole Christ², Julian Gannon¹, Yilun Liu¹, Tim Hunt¹, Maria Jasins² & Stephen C. West¹

¹Cancer Research UK, London Research Institute, Clare Hall Laboratories, South Mimms, Hertfordshire EN6 3LD, UK

²Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

Inherited mutations in *BRCA2* are associated with a predisposition to early-onset breast cancers. The underlying basis of tumorigenesis is thought to be linked to defects in DNA double-strand break repair by homologous recombination. Here we show that the carboxy-terminal region of *BRCA2*, which interacts directly with the essential recombination protein RAD51, contains a site (serine 3291; S3291) that is phosphorylated by cyclin-dependent kinases. Phosphorylation of S3291 is low in S phase when recombination is active, but increases as cells progress towards mitosis. This modification blocks C-terminal interactions between *BRCA2* and RAD51. However, DNA damage overcomes cell cycle regulation by decreasing S3291 phosphorylation and stimulating interactions with RAD51. These results indicate that S3291 phosphorylation might provide a molecular switch to regulate RAD51 recombination activity, providing new insight into why *BRCA2* C-terminal deletions lead to radiation sensitivity and cancer predisposition.

Individuals carrying a mutant allele of *BRCA2* are genetically predisposed to the development of breast, ovarian and other cancers. About 80% of these individuals will develop cancers during their lifetime, and their tumours are characterized by a loss of heterozygosity^{1–3}. The sequence of events that take place from the initial germline mutation to loss of the wild-type allele is not yet clear, but the consequences are catastrophic as indicated by the gross chromosomal instability phenotype associated with *BRCA2*-defective cells grown in culture^{4–6}.

The *BRCA2* gene encodes a large (384-kDa) protein containing 3,418 amino acids^{7,8}. Although the sequence of *BRCA2* reveals few clues to its cellular function, a role in DNA repair and replication fork maintenance is indicated by the hypersensitivity of *BRCA2*-defective cells to DNA-damaging agents^{4,9,10}. Specifically, *BRCA2* mutant cell lines are compromised in their ability to repair DNA double-strand breaks by homologous recombination^{11,12}.

The function of *BRCA2* in homologous recombination is likely to be mediated through interactions with RAD51, a protein that promotes the essential homologous-pairing and strand-exchange phases necessary for the recombinational repair of DNA damage¹³. Within exon 11 of *BRCA2*, a series of BRC repeats that bind RAD51 have been described^{14–17}. An additional, but unrelated, RAD51 interaction domain has been mapped to exon 27 at the C terminus of *BRCA2* (refs 9, 18).

In vitro, BRC peptides bind RAD51 monomers and block formation of the RAD51 nucleoprotein filaments necessary for recombinase activity^{19–21}. Related interactions may occur *in vivo*, because ectopic expression of BRC fragments confers a dominant-negative (recombination-defective) phenotype on wild-type cells^{22,23}. In response to DNA damage, however, *BRCA2* and RAD51 colocalize to nuclear foci where DNA repair reactions are thought to take place^{24–26}. Cell lines lacking *BRCA2* do not elicit this repair response, and RAD51 localization to foci is reduced^{27–29}. These observations led to the proposal that *BRCA2* has a dual function in RAD51 control: by acting as a scaffold for RAD51 binding while enabling the rapid association of RAD51 to sites of DNA damage in response to cellular stress¹⁹. Although this is consistent with the recombinational repair defect and genome instability observed in *BRCA2*-deficient cells, precisely how the

BRC repeats and the C-terminal region of *BRCA2* bind and release RAD51 at the appropriate time and in response to a particular signal is unknown. Here we focus on the highly conserved *BRCA2* C-terminal region that interacts with RAD51, and define its crucial function in homologous recombination.

Phosphorylation of the C terminus of *BRCA2*

A series of nine overlapping glutathione S-transferase (GST) fusion proteins designated B2-1 to B2-9 (ref. 30), spanning the entire coding region of *BRCA2*, were used to define the regions of *BRCA2* that interact with RAD51 (Fig. 1a). Purified GST-*BRCA2* fusion proteins were added to HeLa cell-free extracts, and GST-*BRCA2*-RAD51 complexes were isolated with glutathione beads. RAD51 was found to bind B2-3 (containing BRC repeats 1 and 2), B2-4 (containing BRC3, 4 and 5) and B2-9 (the C-terminal region of *BRCA2*) as detected by western blotting (Fig. 1b, middle panel, lanes 3, 4 and 9). Similar results were obtained with purified RAD51 (rRAD51; Fig. 1b, lower panel). We also determined whether any of the RAD51-interacting fragments were targets for phosphorylation. Incubation of the nine fusions with an extract prepared from asynchronous HeLa cells in the presence of [γ -³²P]ATP revealed that B2-1, B2-2, B2-7 and B2-9 were phosphorylated (Fig. 1c), leading us to further explore the possibility that phosphorylation affects interactions between the *BRCA2* C terminus (B2-9) and RAD51.

First, the kinases(s) responsible for B2-9 phosphorylation were identified. When the extract was fractionated by precipitation with ammonium sulphate, followed by chromatography on phosphocellulose and butyl-sepharose (BS), we found that the activity migrated with histone H1 phosphorylation activity (Fig. 1d). Because histone H1 is a known substrate for cyclin-dependent kinases (CDKs) these results implicated CDKs. Confirmation was obtained when it was shown that B2-9 phosphorylation was blocked by roscovitine, a chemical inhibitor of CDK, and by p21, a cyclin A- and cyclin E-associated CDK inhibitor (Fig. 1e). Similar results were obtained with CDC6, a known protein target for CDK³¹. B2-9 could also be phosphorylated by purified CDK2-cyclin A (Fig. 1f) in a reaction that was dependent on a cyclin-binding sequence, known as a Cy or RXL motif³², as indicated by

inhibition with a Cy peptide inhibitor (Fig. 1g). Control reactions showed that the Cy peptide did not inhibit histone H1 phosphorylation, which is consistent with the lack of a Cy motif in H1.

Division of B2-9 into three smaller GST fusions (designated TR1, TR2 and TR3; Fig. 1a) revealed that the kinase present in butyl-sepharose fraction 15 phosphorylated the overlapping fragments TR1 and TR2, whereas little activity was observed on TR3 (Fig. 1d, e, and data not shown). The phosphorylation of TR2 was particularly interesting because the RAD51 interaction domain maps exclusively to TR2, as determined by pull-down assays (Fig. 2b, lane 3). Sequence analysis of the TR2 region shows that it is highly conserved between mammalian species and contains five conserved S/TP potential CDK target motifs (Fig. 2a, red letters). Consistent with the data of Fig. 1g, the TR2 region also contains a Cy (RXL)-binding motif (amino acids 3,269–3,271; Fig. 2a, green box). Subsequent analyses were therefore limited to the RAD51 interaction domain TR2.

S3291 phosphorylation modulates RAD51 binding

To determine whether the five S/TP potential CDK target motifs in TR2 might be important for regulating interactions with RAD51, a series of glutamate-substituted mutants were constructed that mimic a negatively charged phosphate residue (Fig. 2c, constructs

designated a to m). Strikingly, glutamate substitution at S3291 blocked interactions between B2-9 and RAD51 (Fig. 2d, lanes 3, 4 and 7–13). All other mutants interacted normally with RAD51. Most importantly, we found that phosphorylation of S3291 modulated the ability of TR2 to interact with RAD51. To show this, a series of biotinylated peptides were synthesized that were phosphorylated at defined positions (Fig. 2e). When incubated with RAD51, the three peptides phosphorylated at S3291 did not bind RAD51 (Fig. 2f, lanes 3, 5 and 6).

Dynamic regulation of S3291 phosphorylation

Because these data indicate that phosphorylation of S3291 by CDK might have a crucial function in modulating the ability of BRCA2 to interact with RAD51, we next analysed the *in vivo* phosphorylation status of S3291 with a phospho-specific antibody (S3291Ph). GST-B2-9, overexpressed in asynchronous HT1080 cells, was weakly recognized by S3291Ph (Fig. 3a, lane 3), whereas after nocodazole-induced mitotic arrest we observed a strong interaction with the phospho-specific antibody (Fig. 3a, lane 4). Similar results were obtained with endogenous full-length BRCA2 (Fig. 3b, compare lanes 3 and 4). After treatment with nocodazole, the CDK responsible for S3291 phosphorylation is likely to be CDK1–cyclin B, because only low concentrations of cyclins A and E were detectable in these

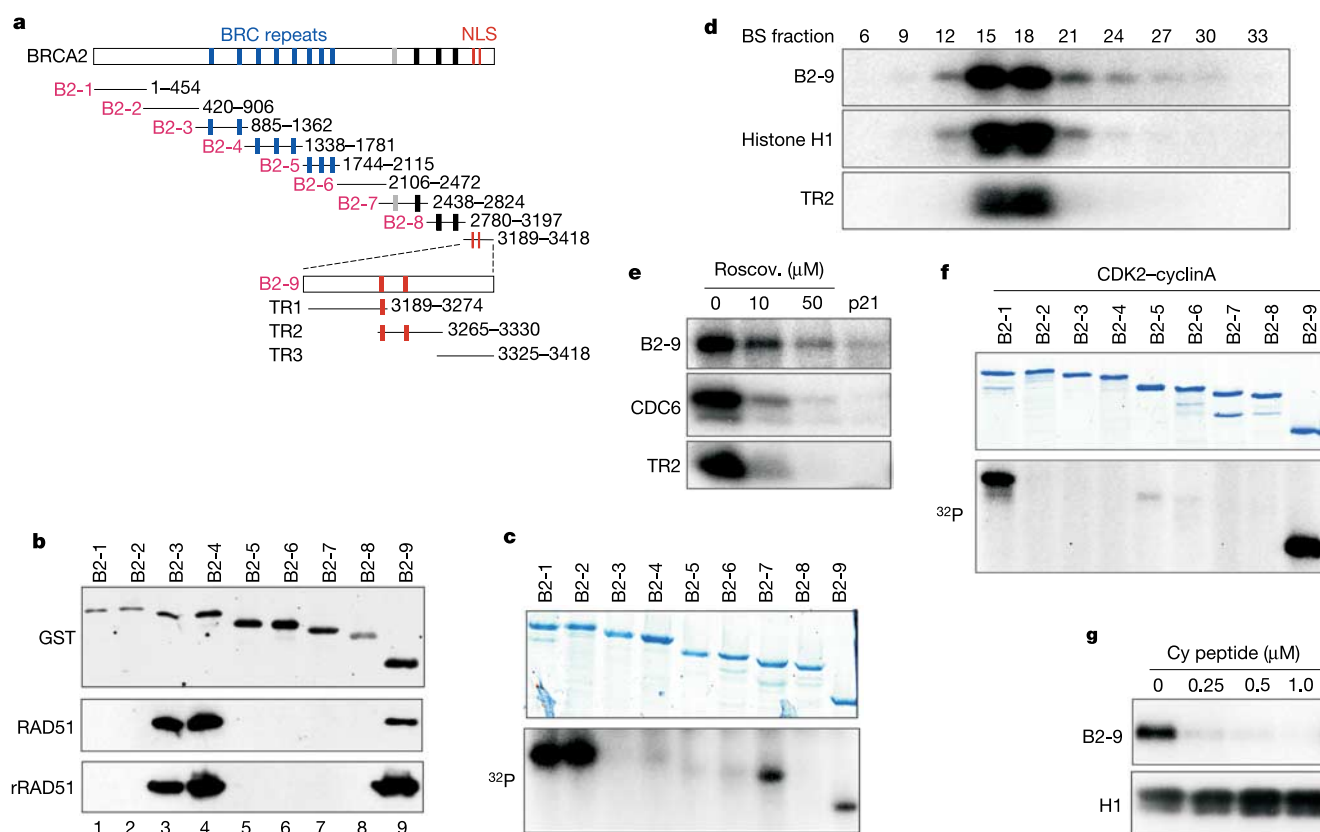


Figure 1 Phosphorylation of the C-terminal RAD51-interaction domain of BRCA2. **a**, Schematic diagram of the nine overlapping GST–BRCA2 fusion proteins, designated B2-1 to B2-9. The B2-9 region was further subdivided into the three overlapping GST fusions designated TR1, TR2 and TR3. The BRC repeats (blue), helical domain (grey), OB folds (black) and NLS sequences (red) are indicated. **b**, Interactions between the GST–BRCA2 fusions B2-1 to B2-9 with RAD51. GST fusion proteins attached to beads were incubated with HeLa extracts or purified recombinant RAD51 (rRAD51). After being washed, the beads were boiled and the GST fusions and bound RAD51 protein were detected by SDS–PAGE followed by western blotting with an anti-GST polyclonal antibody or the anti-RAD51 monoclonal antibody 14B4. **c**, Phosphorylation of B2-1, B2-2, B2-7 and B2-9 by HeLa extracts. GST fusion proteins attached to beads were incubated with an

extract from asynchronous HeLa cells in the presence of $[\gamma\text{-}^{32}\text{P}]\text{ATP}$. After being washed, ^{32}P -labelled products were detected by SDS–PAGE followed by autoradiography. **d**, B2-9 (and TR2) phosphorylation activity co-purifies with the kinase that phosphorylates histone H1. Fractions from the butyl-Sepharose (BS) column are shown. **e**, Inhibition of B2-9 and TR2 phosphorylation by the CDK–cyclin A inhibitors roscovitine and p21. Butyl-Sepharose fraction 15 was used for the kinase assays. CDC6 was a control. **f**, Phosphorylation of the BRCA2 GST fusions by purified CDK2–cyclin A. Reactions were performed as described for **c**, except that purified CDK2–cyclin A (0.7 nM) replaced the HeLa extract. **g**, Dependence of B2-9 phosphorylation on the cyclin A (Cy) recognition sequence. Reactions were performed as described for **f**, except that they were supplemented with a Cy peptide inhibitor. Histone H1 was used as a control.

extracts (Fig. 3c). We did not observe S3291Ph signals when extracts were treated with lambda phosphatase (data not shown).

To determine the phosphorylation status of S3291 throughout the cell cycle, HeLa cells were synchronized at the G1–S boundary using a double thymidine block. Whole cell extracts were prepared at various times after release from cell cycle arrest and probed with the S3291 phospho-specific antibody as well as a variety of cyclin antibodies. In these experiments, the phospho-specific antibody MPM-2 was used to detect mitotic entry. S3291 phosphorylation occurred as the cells progressed from G2 to M phase after the

expression of both cyclin A and cyclin B (Fig. 3d). Kinase activity measurements, determined after immunoprecipitation of cyclin A and cyclin B, and using TR2 and histone H1 as substrates, showed that both CDK–cyclin A and CDK1–cyclin B may phosphorylate the RAD51 interaction domain located at the C terminus of BRCA2. However, the overall phosphorylation status of S3291 is likely to be affected not only by these cyclin-associated kinase activities but also by changes to phosphatase activities and/or factors that modulate substrate recognition by CDKs throughout the cell cycle.

The S3291 phospho-specific antibody was also used to probe

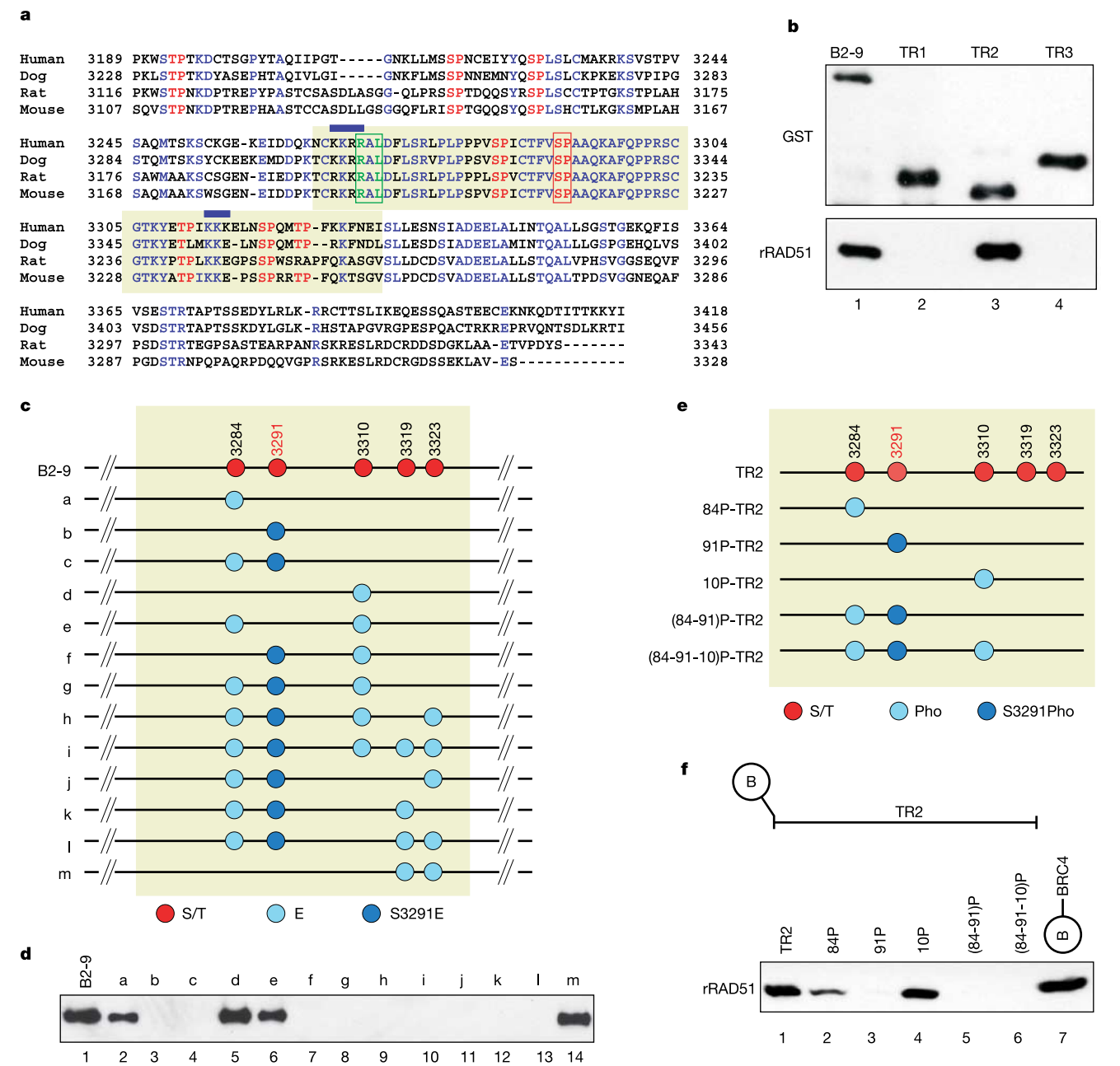


Figure 2 S3291 phosphorylation blocks interactions between the C terminus of BRCA2 and RAD51. **a**, Sequence alignment of the C-terminal (B2-9) region of BRCA2 from different species. The TR2 region (yellow shading) containing two NLS sequences (blue bars) is indicated. Conserved amino acids (blue) and CDK target sites (red letters) are highlighted. The green box indicates the cyclin A recognition motif (Cy or R_xL); the red box indicates the CDK target site at S3291. **b**, Interaction of GST–TR2 with purified RAD51 protein. Reactions were performed as described in Fig. 1b. **c**, Glutamate-substituted B2-9 mutants in which one (or more) CDK consensus phosphorylation site was mutated.

d, Glutamate substitution at S3291 blocks interaction with RAD51. All mutant GST fusion proteins (designated a to m as shown in **c**) were purified and mixed with recombinant RAD51; interactions were detected by pull-down assays. **e**, Synthetic TR2 phosphopeptides in which the various CDK target sites were phosphorylated (Pho). **f**, Phosphorylation of S3291 inhibits the interaction of TR2 with RAD51. Biotinylated (indicated by B in a circle) TR2 phosphopeptides were mixed with recombinant RAD51, and complexes were detected by streptavidin pull-downs. Biotinylated BRC4 peptide was used as a control.

full-length endogenous BRCA2 in extracts made from cells at various times after treatment with ionizing radiation. Phosphorylation of S3291 decreased after ionizing radiation (Fig. 4a), a change that was coupled with a more than 50% decrease in kinase activity (Fig. 4b). A decrease in S3291 phosphorylation was not observed in ATM-deficient cells (data not shown). Inhibition of CDK activity in normal cells by treatment with the CDK inhibitor roscovitine promoted a similar effect to radiation, contrasting with the large increase in kinase activity observed after treatment with nocodazole. Using cells stably expressing GST-B2-9, we found that decreased S3291 phosphorylation caused by ionizing radiation was associated with a doubling in ability to bind RAD51 (Fig. 4c). Thus, S3291

phosphorylation, which is decreased by treatment with ionizing radiation or stimulated by a nocodazole-induced block to cell cycle progression, is inversely correlated with RAD51 binding ability. These results show that the phosphorylation status of S3291 is crucial for modulating interactions between the C terminus of BRCA2 and RAD51 protein.

TR2 expression affects double-strand break repair

Analysis of the Breast Cancer Information Core database revealed that there are four independent examples of mutations in the S3291-P3292 CDK target sequence in individuals identified by increased cancer incidence. We therefore wished to compare the efficiency of recombinational repair in wild-type cells with those carrying an S3291 mutation in BRCA2. Initially, attempts were made to measure homologous recombination efficiency in BRCA2-defective human and hamster cells complemented with either wild-type or S3291 mutant derivatives of BRCA2. However, we found that BRCA2-defective cell lines grew very poorly and exhibited massive cell death after transfection (data not shown). As an alternative approach, we determined whether expression of GST-TR2, or a glutamate-substituted derivative (designated S3291E) that mimics a constitutively phosphorylated form of the protein, exerted different effects on the efficiency of recombinational repair. Preliminary studies showed that GST-TR2, but not the S3291E derivative, when expressed *in vivo*, interacted with RAD51 protein (Fig. 4d, lanes 2 and 4). We therefore used I-SceI endonuclease to induce chromosomal double-strand breaks in cells overexpressing TR2 or S3291E and determined the efficiency of double-strand break repair by the formation of green fluorescent protein (GFP)-positive cells²² (Fig. 4e).

Expression of TR2 reduced the efficiency of recombinational repair by 50%, whereas S3291E did not affect recombination (Fig. 4f). Control experiments, in which I-SceI cleavage was not induced, indicated that cell viability was unaffected by expression of the TR2 constructs (data not shown). These results show that overexpression of the non-phosphorylated RAD51-interaction domain of BRCA2 results in a dominant-negative recombination-deficient phenotype, whereas expression of the S3291E derivative had no effect on recombination efficiency. Additional experiments showed that expression of a construct carrying a serine → alanine mutation at S3291 also did not confer a dominant-negative phenotype (Fig. 4f). As with S3291E, the S3291A derivative did not bind RAD51 (Fig. 4c). These studies show that interactions between the C-terminal region of endogenous BRCA2 and RAD51 are necessary for efficient double-strand break repair by homologous recombination.

In control experiments, we found that a peptide 38 amino acids long (residues 3,265–3,302) interacts with RAD51, whereas an alanine or glutamate substitution at S3291 blocks the interaction. In contrast, conversion of Pro 3292 (P3292) to leucine did not affect RAD51 binding (data not shown). Because it is unlikely that the S3291A derivative would cause a conformational change in a small peptide but the P3292 derivative would not, these experiments exclude the possibility that the observed effects are due to gross conformational changes and emphasize that S3291 is a crucial residue necessary for interaction with RAD51.

A mechanism for RAD51 control

The work described here provides new insight in the involvement of BRCA2 in break repair by homologous recombination. Most importantly, the phosphorylation status of S3291 seems crucial for regulating the interaction of RAD51 with the C-terminal region of BRCA2. We found that radiation treatment resulted in reduced CDK-mediated phosphorylation of S3291, which stimulated the interaction between the BRCA2 C terminus and RAD51. Moreover, during the cell cycle, we observed that S3291 phosphorylation increased throughout G2/M and was reduced in G1. The low levels

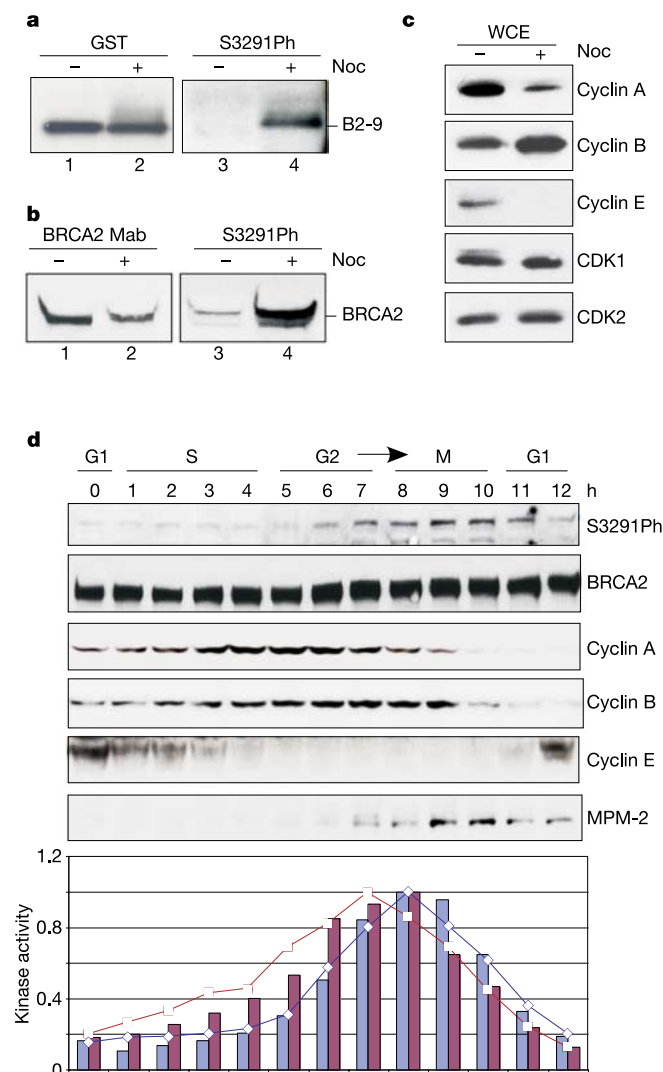


Figure 3 Analysis of the phosphorylation status of S3291 *in vivo*. **a**, GST-B2-9 was overexpressed in HT1080 cells. The cells were treated with or without nocodazole (Noc), and after 24 h extracts were prepared and GST-B2-9 was pulled down with the GST tag. Samples analysed by SDS-PAGE were then probed with an anti-GST antibody or the phospho-specific S3291 antibody. **b**, HeLa cells were treated with or without nocodazole and extracts probed for full-length BRCA2 using a mouse monoclonal BRCA2 antibody (OP95) or the S3291 phospho-specific antibody. **c**, Whole extracts made from cells treated with or without nocodazole were blotted with the indicated monoclonal antibodies. **d**, HeLa cells were synchronized by using a double thymidine block. At various times after release samples were taken; extracts were prepared and blotted with the indicated antibodies. In the lower panel, cyclin A (red) or cyclin B (blue) immunoprecipitates were analysed for kinase activity with TR2 (lines) and histone H1 (histograms) as substrates. Maximum kinase activity was arbitrarily designated as 1.0.

of phosphorylation observed at S phase are likely to be important for the repair of replication-induced breaks. Conversely, phosphorylation of S3291 in mitosis might be responsible for the inactivation of homologous recombination, and/or might contribute to the degradation of BRCA2 before BRCA2 synthesis occurs *de novo* in late G1 (ref. 33).

The results provide new insight into a mechanism by which checkpoint control and recombinational repair are coupled in mammalian cells. The repair pathway is thought to involve ATM, p53 and CHK2, all of which are related to breast cancer susceptibility and are involved in CDK inactivation and cell cycle arrest after DNA damage. ATM transduces the DNA damage signal to p53, leading to the transcriptional activation of p21 that inhibits CDK activity³⁴. ATM also phosphorylates and activates CHK2, which in turn phosphorylates CDC25 phosphatase, leading to its degradation or cytoplasmic sequestration³⁵. Because CDC25 is required for CDK activation, loss of CDC25 activity will help to maintain the phosphorylated/inactivated state of CDK³⁶. Inactivation of CDK will reduce its ability to phosphorylate S3291 of BRCA2, thus stimulating interactions between RAD51 and the C-terminal region of BRCA2.

The consequences of loss of phosphorylation of S3291 after CDK inhibition lead us to suggest a model in which the BRC repeats and the C-terminal RAD51 interaction domain are functionally distinct. As suggested previously, under normal conditions interactions at the BRC repeats may serve to inactivate RAD51 by providing a scaffold for monomer binding^{19,20}. It is expected that phosphorylation of S3291 by CDK will then block interactions between RAD51 and the C terminus of BRCA2. However, after ionizing radiation, and possibly other DNA-damaging treatments, decreased CDK activity leads to the loss of phosphorylation at S3291 and activation of the C-terminal RAD51 interaction domain.

Structural studies of the C-terminal region of BRCA2 complexed with DSS1 protein revealed the presence of three globular domains that form OB folds, which are also found in single-strand DNA-binding proteins such as RPA³⁷. It is therefore tempting to speculate that the binding of RAD51 by the C-terminal interaction domain of BRCA2, which occurs after the loss of S3291 phosphorylation, might contribute to the loading of RAD51 to single-stranded DNA. The binding of RAD51 by this region might be an important component of the cellular DNA damage response by accepting RAD51 monomers from the BRC repeats and facilitating their

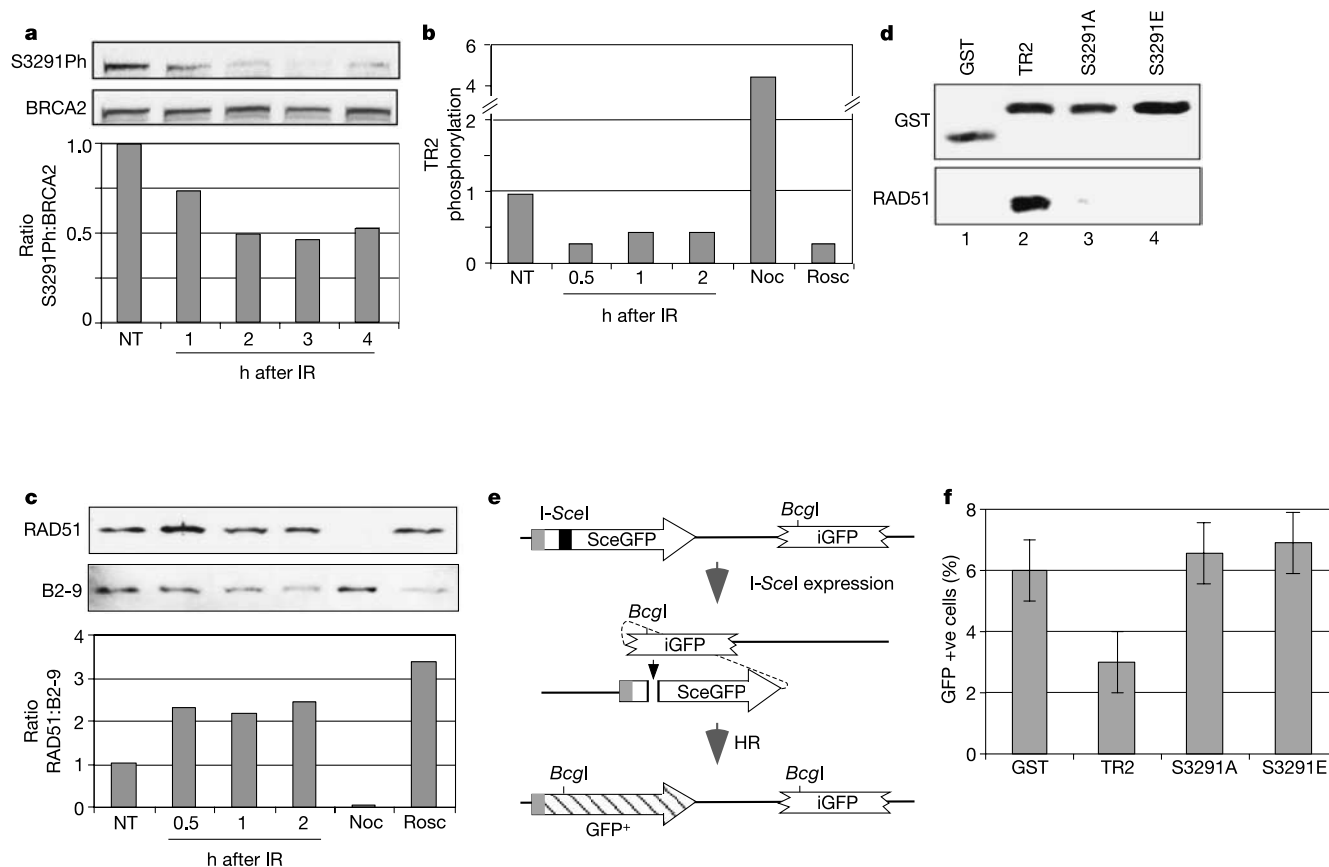


Figure 4 Involvement of S3291 phosphorylation in mediating RAD51 interactions and efficient homologous recombination. **a**, HT1080 cells were treated with 4 Gy ionizing radiation and at the indicated times whole cell extracts were prepared and blotted with S3291 phospho-specific antibody or the BRCA2 monoclonal antibody OP95. Only the section of the blot showing full-size BRCA2 is shown. The ratio of BRCA2 phosphorylated at S3291 to unphosphorylated BRCA2 is indicated. **b**, Phosphorylation of TR2 after treatment with ionizing radiation (IR) or drugs. **c**, HT1080 cells expressing GST-B2-9 fusion protein were left untreated (NT) or treated with 4 Gy ionizing radiation, 0.13 μ M nocodazole (Noc) or 10 μ M roscovitine (Rosco). Irradiated cells were then grown for the indicated durations. Drug treatment was maintained for 24 h, at which time extracts were prepared and GST pull-down assays were used to determine the amounts of RAD51 associated with B2-9. **d**, Wild-type TR2, S3291A or S3291E, were expressed as GST

fusions in 293T cells, and interactions with endogenous RAD51 were detected using GST pull-down assays. **e**, Schematic diagram indicating the double-strand break repair assay. 293DR-GFP cells have two GFP fragments integrated into genomic DNA, one of which has a unique I-SceI cleavage site. Expression of I-SceI endonuclease introduces a double-strand break that can be repaired by homologous recombination (HR) to produce a functional GFP gene (hatched). **f**, GST-TR2, GST-S3291A or GST-S3291E expression vectors were co-transfected with the I-SceI expression vector into 293DR-GFP cells. Recombinational repair of the induced double-strand break was analysed by fluorescence-activated cell sorting analysis. The error bars show standard deviations for three independent experiments ($n = 3$). A highly significant P value (equivalent to 0.000769) relating to I-SceI/GST and I-SceI/TR2 was determined by using a standard t -test.

transfer to DNA, where functional RAD51 nucleoprotein filaments are assembled. Indeed, it is possible that the C-terminal region of BRCA2 could fulfil a key role in the hand-over of RAD51 to the initiating DNA substrate.

The involvement of CDK in the recruitment of RAD51 to DNA break sites was recently shown in *Saccharomyces cerevisiae*^{38,39}. These studies indicated that CDKs in yeast promote the activation of homologous recombination, in contrast to their role in mammalian cells. This difference might be due the fact that, in response to DNA damage, budding yeasts arrest their cell cycle in the mitotic phase with elevated CDK activity, rather than in G2 as in other organisms. Moreover, a functional BRCA2 homologue has not been identified in yeast.

Although it is unlikely that S3291 phosphorylation provides the only regulatory mechanism that controls recombinase activity, this model provides a simple explanation for the activation of RAD51 in response to DNA-damaging agents. It is also consistent with observations showing that the deletion of exon 27 results in recombination deficiency and radiation sensitivity^{12,40–42}. Moreover, exon 27 deletion, or a more extensive deletion of the C-terminal half of BRCA2 that leaves several of the BRC repeats intact, results in a loss of ionizing-radiation-induced RAD51 focus formation^{40,41}. The importance of such dynamic changes to the S3291 phosphorylation site is also highlighted by observations showing that it is a cancer-related mutation site, as indicated by the incidence of P3292L mutations in individuals affected with breast cancer. □

Methods

Cell culture

Human 293T and HT1080 cells were cultured on plates in DMEM medium supplemented with 10% v/v fetal bovine serum and streptomycin/penicillin (100 units ml⁻¹). For 293DR-GFP cells the plates were coated with 0.01% poly-(L-lysine). For the thymidine block, 25% confluent HeLa cells were cultured in the presence of 2 mM thymidine for 19 h, and released into fresh medium for 9 h. They were then incubated in the presence of 2 mM thymidine for 16 h before release into fresh medium.

Plasmids

The GST constructs B2-1 to B2-9 have been described³⁰. GST-TR1, GST-TR2 and GST-TR3 were constructed by cloning by polymerase chain reaction (PCR) into pGEX-4T3 (Amersham) and the mammalian expression vectors pGST-B2-9, pGST-TR1, pGST-TR2 and pGST-TR3 were made by PCR cloning into the Gateway entry vector pDONR221 and transferred into pDEST27 (Invitrogen).

Proteins

Recombinant RAD51, CDK2-cyclin A and GST-p21N were prepared as described^{43–45}. The GST-p53 peptide was purified from pGEX 2T p53(9–22), a gift from Dr M. Kastan. CDC6 (residues 1–110) was purified as a GST fusion protein. Histone H1 was purchased from Upstate. The C-peptide PSACRNLFPGVD corresponds to residues 26–37 of p27. All peptides were prepared by solid-phase synthesis and purified by high-performance liquid chromatography; their sequences were confirmed by mass spectroscopy. Where indicated, peptides contained an N-terminal biotin group with an aminohexanoic acid spacer.

BRCA2 GST fusion plasmids were grown in BL21 RIL codon plus *Escherichia coli* (Stratagene) and protein was expressed by incubation overnight at 18 °C after the addition of 10 μM isopropyl β-D-thiogalactoside. Proteins were purified from soluble extracts with glutathione Sepharose 4 fast flow beads (Amersham).

Mammalian extracts

Mammalian cell pellets (about 0.6 g) were suspended in 3 ml HE buffer (20 mM Hepes-NaOH pH 7.6, 2 mM EGTA, 1.5 mM MgCl₂, 50 mM NaF, 0.5% Nonidet P40, 10% glycerol, 1 mM Na₂VO₄, 20 mM β-glycerophosphate, 1 mM dithiothreitol, 1.5 μg ml⁻¹ aprotinin and 1 μg ml⁻¹ each of leupeptin, pepstatin A and chymostatin) containing 450 mM KCl, and placed on ice for 30 min. After centrifugation in a Microfuge for 10 min, the supernatant was diluted with 2 volumes of HE buffer (to 150 mM KCl), and then centrifuged at 35,000 r.p.m. for 60 min in a Beckman 70.1Ti rotor. Supernatants were used for protein interaction studies and kinase assays.

In vitro protein interactions

GST fusion proteins (about 1 μg) bound to glutathione beads were mixed with HeLa extract (1 ml; about 5 mg) and incubated for 2.5 h at 4 °C. Alternatively, the bead-bound fusions were mixed with BSA (5 mg ml⁻¹) in HE buffer supplemented with 150 mM KCl, and after 30 min they were supplemented with RAD51 (50 ng). After a further 2 h, the beads were washed with HE buffer containing 150 mM KCl. Complexes were then boiled in SDS sample buffer and analysed by SDS-polyacrylamide-gel electrophoresis (SDS-PAGE) followed by western blotting.

Interactions between biotinylated peptides (100 ng) and RAD51 (80 ng) were

performed similarly, and complexes were analysed by pull-down assays with streptavidin beads, followed by SDS-PAGE and western blotting.

Kinase assays

Protein substrates (2 μg in 20 μl) were phosphorylated in kinase buffer (10 mM Hepes-HCl pH 7.6, 50 mM β-glycerophosphate, 50 mM NaCl, 10 mM MgCl₂, 10 mM MnCl₂, 5 μM ATP, 1 mM dithiothreitol, 1 μCi [γ-³²P]ATP) by the addition of 1 μl cell extract or fractionated extract. After 30 min at 30 °C, reactions were stopped by the addition of SDS sample buffer and boiled for 5 min. Proteins were analysed by SDS-PAGE and revealed by Coomassie blue staining and autoradiography.

Generation of 293DR-GFP reporter cells

About 5 × 10⁶ 293 Tet-Off cells (Clontech) were transfected with 50 μg *SacI/KpnI*-linearized hprtDR-GFP reporter vector⁴⁶ by electroporation. After 24 h, puromycin selection (1 μg ml⁻¹) was applied and a clone with an intact copy of the reporter (TOS A4) was identified by Southern blotting⁴⁷.

I-SceI-induced double-strand break repair

To examine recombination induced by double-strand breakage, 293DR-GFP cells were co-transfected with the I-SceI expression vector pCBASce⁴⁸ and pGST, pGST-TR2, pGST-TR2A or pGST-TR2E. For transfection, 0.6 ml of cells at 8 × 10⁶ ml⁻¹ were electroporated at 280 V and 960 μF with 30 μg pCBASce and 30 μg BRCA2-derived expression vectors. To determine transfection efficiency, 293DR-GFP cells were transfected under the same conditions with 3 μg GFP-expressing vector pNZE-CAG⁴⁷ together with 27 μg mock plasmid DNA and the BRCA2-derived expression vector (30 μg). For each transfection a single-cell suspension was prepared 48 h after electroporation, and the percentage of GFP-positive cells was determined by flow cytometry on a Becton Dickinson FACScan. Dead cells were excluded from the analysis by staining with 7-amino-actinomycin D (Sigma), and fluorescence-activated cell sorting data were analysed with CellQuest software.

Received 5 November 2004; accepted 27 January 2005; doi:10.1038/nature03404.

- Welsh, P. L. & King, M. C. *BRCA1* and *BRCA2* and the genetics of breast and ovarian cancer. *Hum. Mol. Genet.* **10**, 705–713 (2001).
- Venkitaraman, A. R. Cancer susceptibility and the functions of *BRCA1* and *BRCA2*. *Cell* **108**, 171–182 (2002).
- King, M. C., Marks, J. H. & Mandell, J. B. Breast and ovarian cancer risks due to inherited mutations in *BRCA1* and *BRCA2*. *Science* **302**, 643–646 (2003).
- Patel, K. J. *et al.* Involvement of *BRCA2* in DNA repair. *Mol. Cell* **1**, 347–357 (1998).
- Tutt, A. *et al.* Absence of *BRCA2* causes genome instability by chromosome breakage and loss associated with centrosome amplification. *Curr. Biol.* **9**, 1107–1110 (1999).
- Yu, V. P. C. *et al.* Gross chromosomal rearrangements and genetic exchange between non-homologous chromosomes following *BRCA2* inactivation. *Genes Dev.* **14**, 1400–1406 (2000).
- Wooster, R. *et al.* Identification of the breast cancer susceptibility gene *BRCA2*. *Nature* **378**, 789–792 (1995).
- Tavtigian, S. V. *et al.* The complete *BRCA2* gene and mutations in chromosome 13q-linked kindreds. *Nature Genet.* **12**, 333–337 (1996).
- Sharan, S. K. *et al.* Embryonic lethality and radiation hypersensitivity mediated by *RAD51* in mice lacking *BRCA2*. *Nature* **386**, 804–810 (1997).
- Lomonosov, M., Anand, S., Sangrithi, M., Davies, R. & Venkitaraman, A. R. Stabilization of stalled DNA replication forks by the *BRCA2* breast cancer susceptibility protein. *Genes Dev.* **17**, 3017–3022 (2003).
- Connor, F. *et al.* Tumorigenesis and a DNA-repair defect in mice with a truncating *BRCA2* mutation. *Nature Genet.* **17**, 423–430 (1997).
- Moynahan, M. E., Pierce, A. J. & Jasin, M. *BRCA2* is required for homology-directed repair of chromosomal breaks. *Mol. Cell* **7**, 263–272 (2001).
- West, S. C. Molecular views of recombination proteins and their control. *Nature Rev. Mol. Cell Biol.* **4**, 1–11 (2003).
- Bork, P., Blomberg, N. & Nilges, M. Internal repeats in the *BRCA2* protein sequence. *Nature Genet.* **13**, 22–23 (1996).
- Bignell, G., Micklem, G., Stratton, M. R., Ashworth, A. & Wooster, R. The BRC repeats are conserved in mammalian *BRCA2* proteins. *Hum. Mol. Genet.* **6**, 53–58 (1997).
- Wong, A. K. C., Pero, R., Ormonde, P. A., Tavtigian, S. V. & Bartel, P. L. *RAD51* interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene *BRCA2*. *J. Biol. Chem.* **272**, 31941–31944 (1997).
- Chen, P. L. *et al.* The BRC repeats in *BRCA2* are critical for *RAD51* binding and resistance to methyl methanesulphonate treatment. *Proc. Natl Acad. Sci. USA* **95**, 5287–5292 (1998).
- Mizuta, R. *et al.* Rab22 and Rab163/mouse *BRCA2*: proteins that specifically interact with the *RAD51* protein. *Proc. Natl Acad. Sci. USA* **94**, 6927–6932 (1997).
- Davies, A. A. *et al.* Role of *BRCA2* in control of the *RAD51* recombination and DNA repair protein. *Mol. Cell* **7**, 273–282 (2001).
- Pellegrini, L. *et al.* Insights into DNA recombination from the structure of a *RAD51*-*BRCA2* complex. *Nature* **420**, 287–293 (2002).
- Pellegrini, L. & Venkitaraman, A. Emerging functions of *BRCA2* in DNA recombination. *Trends Biochem. Sci.* **29**, 310–316 (2004).
- Stark, J. M. *et al.* ATP hydrolysis by mammalian *RAD51* has a key role during homology-directed DNA repair. *J. Biol. Chem.* **277**, 20185–20194 (2002).
- Chen, C. F., Chen, P. L., Zhong, Q., Sharp, Z. D. & Lee, W. H. Expression of BRC repeats in breast cancer cells disrupts the *BRCA2*-*RAD51* complex and leads to radiation hypersensitivity and loss of G₂/M checkpoint control. *J. Biol. Chem.* **274**, 32931–32935 (1999).
- Marmorstein, L. Y., Ouchi, T. & Aaronson, S. A. The *BRCA2* gene product functionally interacts with p53 and *RAD51*. *Proc. Natl Acad. Sci. USA* **95**, 13869–13874 (1998).
- Chen, J. J. *et al.* Stable interaction between the products of the *BRCA1* and *BRCA2* tumor suppressor genes in mitotic and meiotic cells. *Mol. Cell* **2**, 317–328 (1998).

26. Tarsounas, M., Davies, A. A. & West, S. C. RAD51 localization and activation following DNA damage. *Phil. Trans. R. Soc. Lond. B* **359**, 87–93 (2004).
27. Yuan, S.-S. F. *et al.* BRCA2 is required for ionizing radiation-induced assembly of RAD51 complex *in vivo*. *Cancer Res.* **59**, 3547–3551 (1999).
28. Godthelp, B. C., Artwert, F., Joenje, H. & Zdzienicka, M. Z. Impaired DNA damage-induced nuclear RAD51 foci formation uniquely characterizes Fanconi anemia group D1. *Oncogene* **21**, 5002–5005 (2002).
29. Tarsounas, M., Davies, D. & West, S. C. BRCA2-dependent and independent formation of RAD51 nuclear foci. *Oncogene* **22**, 1115–1123 (2003).
30. Lee, M., Daniels, M. J. & Venkitaraman, A. R. Phosphorylation of BRCA2 by the Polo-like kinase Plk1 is regulated by DNA damage and mitotic progression. *Oncogene* **23**, 865–872 (2004).
31. Petersen, B. O., Lukas, J., Sorensen, C. S., Bartek, J. & Helin, K. Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization. *EMBO J.* **18**, 396–410 (1999).
32. Wohlschlegel, J. A., Dwyer, B. T., Takeda, D. Y. & Dutta, A. Mutational analysis of the Cy motif from p21 reveals sequence degeneracy and specificity for different cyclin-dependent kinases. *Mol. Cell. Biol.* **21**, 4868–4874 (2001).
33. Bertwistle, D. *et al.* Nuclear location and cell cycle regulation of the BRCA2 protein. *Cancer Res.* **57**, 5485–5488 (1997).
34. Brugarolas, J. *et al.* Radiation-induced cell cycle arrest compromised by p21 deficiency. *Nature* **12**, 552–557 (1995).
35. Falck, J., Mailand, N., Syljuasen, R. G., Bartek, J. & Lukas, J. The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. *Nature* **410**, 842–847 (2001).
36. Jinno, S. *et al.* Cdc25 is a novel phosphatase functioning early in the cell cycle. *EMBO J.* **13**, 1549–1556 (1994).
37. Yang, H. J. *et al.* BRCA2 function in DNA binding and recombination from a BRCA2–DSS1–ssDNA structure. *Science* **297**, 1837–1848 (2002).
38. Ira, G. *et al.* DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1. *Nature* **431**, 1011–1016 (2004).
39. Aylon, Y., Liefshitz, B. & Kupiec, M. The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle. *EMBO J.* **23**, 4868–4875 (2004).
40. Tutt, A. *et al.* Mutation in BRCA2 stimulates error-prone homology-directed repair of DNA double-strand breaks occurring between repeated sequences. *EMBO J.* **20**, 4704–4716 (2001).
41. Tutt, A. *et al.* Cell cycle and genetic background dependence of the effect of loss of BRCA2 on ionizing radiation sensitivity. *Oncogene* **22**, 2926–2931 (2003).
42. Donoho, G. *et al.* Deletion of BRCA2 exon 27 causes hypersensitivity to DNA crosslinks, chromosomal instability, and reduced life span in mice. *Genes Chromosom. Cancer* **36**, 317–331 (2003).
43. Baumann, P., Benson, F. E., Hajibaghery, N. & West, S. C. Purification of human RAD51 protein by selective spermidine precipitation. *Mutat. Res. DNA Repair* **384**, 65–72 (1997).
44. Brown, N. R. *et al.* The crystal structure of cyclin A. *Structure* **3**, 1235–1247 (1995).
45. Moore, J. D., Kirk, J. A. & Hunt, T. Unmasking the S-phase-promoting potential of cyclin B1. *Science* **300**, 987–990 (2003).
46. Pierce, A. J., Hu, P., Han, M. G., Ellis, N. & Jasin, M. Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells. *Genes Dev.* **15**, 3237–3242 (2001).
47. Pierce, A. J., Johnson, R. D., Thompson, L. H. & Jasin, M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev.* **13**, 2633–2638 (1999).
48. Richardson, C., Moynahan, M. E. & Jasin, M. Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. *Genes Dev.* **12**, 3831–3842 (1998).

Acknowledgements We thank A. Venkitaraman for providing GST–B2-1 to B2-9. F.E. and N.C. are recipients of postdoctoral fellowships from the Human Frontier Scientific Program, and Y.L. is a fellow of the American Cancer Society. This work was supported by Cancer Research UK, the Swiss Bridge Fund and the Breast Cancer Campaign (S.C.W.), and by the Emerald Foundation and an NIH grant (M.J.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.W. (stephen.west@cancer.org.uk).

Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage

Jacob Falck, Julia Coates & Stephen P. Jackson

The Wellcome Trust and Cancer Research UK Gurdon Institute, and Department of Zoology, Cambridge University, Tennis Court Road, Cambridge CB2 1QN, UK

Ataxia-telangiectasia mutated (ATM), ataxia-telangiectasia and Rad3-related (ATR) and DNA-dependent protein kinase catalytic subunit (DNA-PKcs) are members of the phosphoinositide-3-kinase-related protein kinase (PIKK) family, and are rapidly activated in response to DNA damage. ATM and DNA-PKcs respond mainly to DNA double-strand breaks, whereas ATR is activated by single-stranded DNA and stalled DNA replication forks. In all cases, activation involves their recruitment to the sites of damage. Here we identify related, conserved carboxy-terminal motifs in human Nbs1, ATRIP and Ku80 proteins that are required for their interaction with ATM, ATR and DNA-PKcs, respectively. These motifs are essential not only for efficient recruitment of ATM, ATR and DNA-PKcs to sites of damage, but are also critical for ATM-, ATR- and DNA-PKcs-mediated signalling events that trigger cell cycle checkpoints and DNA repair. Our findings reveal that recruitment of these PIKKs to DNA lesions occurs by common mechanisms through an evolutionarily conserved motif, and provide direct evidence that PIKK recruitment is required for PIKK-dependent DNA-damage signalling.

DNA damage response pathways involve three major groups of evolutionarily conserved proteins that act together to translate the signal of damaged DNA into responses of cell cycle arrest and DNA repair. These groups comprise sensor proteins that recognize damaged DNA directly or indirectly, transducer proteins—typically PIKKs such as ATM, ATR and DNA-PKcs—that relay and amplify the damage signal, and effector proteins that control cell cycle progression, chromatin restructuring and DNA repair^{1–3}.

Recruitment of DNA damage-associated PIKKs to DNA lesions is thought to be a principal step in their activation and in their function in checkpoint signalling and DNA repair^{4,5}. Although these PIKKs have an affinity for DNA^{6–8}, recruitment to DNA lesions is facilitated by specific partner proteins. For example, recruitment of ATR to single-stranded DNA (ssDNA) requires ATRIP^{9,10}, and DNA-PKcs recruitment to DNA double-strand breaks (DSBs) is mediated by the Ku70–Ku80 heterodimer^{11,12}. Recently, the evolutionarily conserved Mre11–Rad50–Nbs1 (MRN) protein complex was implicated in ATM recruitment to DSBs^{13–15}.

Here, we describe a conserved region in the C terminus of Nbs1 that serves to recruit activated ATM to sites of DNA damage, thus promoting the phosphorylation of ATM targets and ensuing events of the DNA damage response. Furthermore, we reveal that regions with sequence homology to the Nbs1 C terminus exist in the C-terminal regions of ATRIP and Ku80, and establish that these regions are necessary for these factors to interact with ATR and DNA-PKcs, respectively. Collectively, our findings reveal that these PIKK interaction partners function as DNA damage-specific recruitment modules *in vivo*, and that recruitment of PIKKs to DNA damage is a requisite for PIKK-dependent signalling.

The C terminus of Nbs1 recruits ATM to DSBs

In addition to the previously characterized regions of Nbs1 (Fig. 1a; see also ref. 16), we noted a conserved motif at its extreme C terminus. We reasoned that this might constitute a protein interaction motif, and therefore used a synthetic peptide corresponding to this region to retrieve potential interacting factors. Mass spectrometry analysis of the isolated proteins revealed that ATM

bound to the peptide (Supplementary Fig. S1). Subsequent experiments confirmed that the C-terminal 20 residues of human Nbs1 were sufficient for interaction with ATM, whereas a peptide corresponding to the Mre11 interaction region of Nbs1 (MIR) did not interact with ATM (Fig. 1b). The Nbs1 C terminus also bound directly to highly purified recombinant ATM (Supplementary Fig. S1). Consistent with the sequence conservation between the C termini of Nbs1 orthologues, we found that a peptide corresponding to residues 834–854 of budding yeast Xrs2 interacted with ATM (Fig. 1b).

When expressed in cells, wild-type Nbs1 associated with ATM, but a truncated derivative lacking the C-terminal 20 residues (Nbs1ΔC) did not (Fig. 1c, d). By contrast, Nbs1ΔC still interacted efficiently with the other components of the MRN complex (that is, Mre11 and Rad50; Fig. 1d). An Nbs1 mutant incapable of binding to Mre11 (ΔMIR) still interacted with ATM, albeit less efficiently than wild-type Nbs1 (Fig. 1e), revealing that an intact MRN complex is not essential for the Nbs1–ATM interaction. Closer examination of the Nbs1 C terminus revealed that the residues most conserved throughout evolution were important for the Nbs1–ATM interaction: mutation of residues 736–737 (EE), 741–742 (DD), or 745–746 (RY) to alanine doublets lead in each case to markedly decreased ATM binding (Fig. 1f). Thus, the evolutionarily conserved Nbs1 C terminus constitutes an ATM interaction motif.

To test the function of the Nbs1 C terminus, extracts from complemented NBS fibroblasts were incubated with a double-stranded (ds) DNA fragment coupled to magnetic beads. Notably, ATM only associated efficiently with dsDNA in extracts from wild-type-complemented cells (Fig. 2a). By contrast, binding of Mre11 and Rad50 to the DNA fragments was not affected by Nbs1 status, and wild-type Nbs1 and Nbs1ΔC were detected in similar amounts in DNA pull downs. Thus, the Nbs1 C terminus is needed for efficient recruitment of ATM to DNA ends by MRN. Consistent with this, we found that the association of ATM with DNA fragments was disrupted by a peptide corresponding to Nbs1 residues 734–754 (AIR; Fig. 2b). Ionizing radiation treatment did not enhance complex formation between ATM and Nbs1 in

soluble extracts (Fig. 2c); however, addition of a 30-mer dsDNA oligonucleotide, but not circular DNA, to extracts from untreated cells led to an increased association between Nbs1 and ATM (Fig. 2d), suggesting that MRN binding to DNA ends stabilizes the Nbs1–ATM interaction. In line with these data, an ionizing-radiation-induced increase of chromatin-bound ATM (fraction III) was readily detectable in wild-type- but not vector- or ΔC -complemented NBS cells (Fig. 2e). By contrast, deletion of the ATM interaction motif from Nbs1 did not affect chromatin retention of Nbs1 itself (Fig. 2e).

ATM activation involves its autophosphorylation on Ser1981, and this autophosphorylated species forms ionizing-radiation-induced foci that co-localize with the phosphorylated form of histone H2AX, γ H2AX^{14,15,17}. After irradiation, phospho-S1981-ATM foci that co-localized with γ H2AX and Mre11 were found in NBS cells complemented with wild-type Nbs1 (Fig. 2f). By contrast, phospho-S1981-ATM did not form foci in irradiated vector- or Nbs1 ΔC -complemented cells; instead, increased pan-nuclear staining of phospho-S1981-ATM was observed despite the presence of both γ H2AX and Mre11 foci (Fig. 2f). Importantly, focus formation of Nbs1 and Mre11 was comparable in wild-type- and Nbs1 ΔC -complemented NBS cells (Fig. 2f; see also Supplementary Fig. S2). Together, these data show that recruitment of ATM to sites of DNA damage requires the ATM interaction motif at the Nbs1 C terminus.

The Nbs1–ATM interaction mediates ATM checkpoint functions

Phosphorylation of SMC1, Nbs1, Chk2 and histone H2AX by ATM occurs at sites of DNA damage, whereas phosphorylation of p53, and ATM autophosphorylation itself, are thought to take place throughout the nucleoplasm^{15,18}. To test whether ablation of the ATM–Nbs1 interaction selectively affects phosphorylation of substrates targeted at DSBs, we exposed complemented NBS cells to ionizing radiation and collected them at various time points. In agreement with previous work^{14,15,19,20}, we observed reduced ATM activation in vector-complemented cells as compared with wild-type-complemented NBS cells (Fig. 3a); however, ATM autophosphorylation occurred with similar timing and extent in wild-type- and Nbs1 ΔC -complemented NBS cells (Fig. 3a). Furthermore, Nbs1 status did not affect phosphorylation of p53 on Ser 15 (Fig. 3a). By contrast, phosphorylation of Chk2 and SMC1 on Thr 68 and Ser 966, respectively, was diminished to a similar degree in vector- and Nbs1 ΔC -complemented cells, as compared with wild-type-complemented cells (Fig. 3a). Notably, phosphorylation of Nbs1 on Ser 343 was only detectable in wild-type-complemented cells (Fig. 3a). Interestingly, phosphorylation of histone H2AX was not markedly affected by Nbs1 status (Fig. 3a). To investigate this in more detail, we treated cells with specific inhibitors of either DNA-PK or ATM^{21,22}. This revealed that the apparent independence of histone H2AX phosphorylation on Nbs1 status reflected an increased contribution by DNA-PK in cells

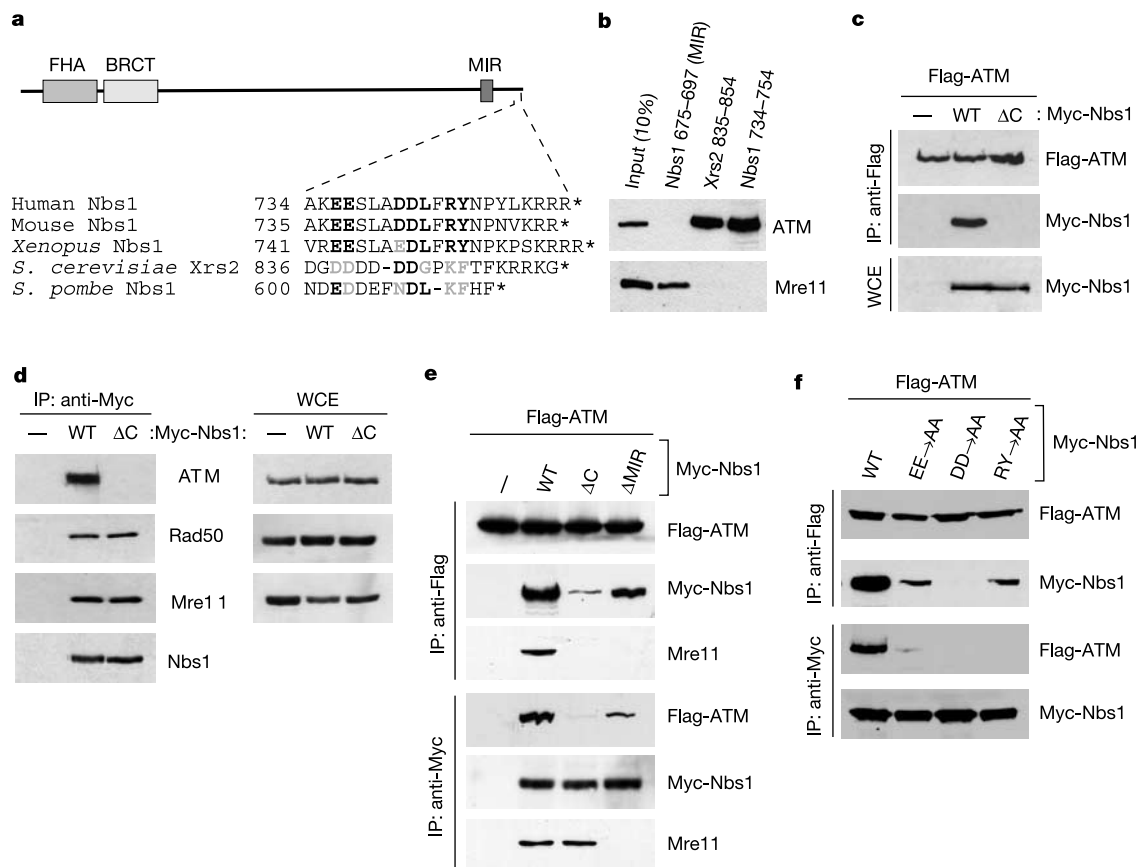


Figure 1 The extreme C terminus of Nbs1 constitutes a conserved ATM interaction motif. **a**, Human Nbs1 with its conserved domains (top) and alignment of the extreme C termini of Nbs1 orthologues from the indicated species (bottom). MIR, Mre11 interaction region. **b**, Biotinylated peptides corresponding to the indicated regions of human Nbs1 and budding yeast Xrs2 were incubated with nuclear extract, isolated and immunoblotted as indicated. **c**, **d**, Extracts from cells expressing Flag-ATM and wild type (WT) or ΔC Myc-

Nbs1 were immunoblotted either directly (WCE, whole cell extract) or after immunoprecipitation as indicated. **e**, **f**, Extracts from cells expressing Flag-ATM and wild type, ΔC , N684A/F685A/K686A/K687A (Δ MIR), E736A/E737A (EE $\rightarrow$ AA), D741A/D742A (DD $\rightarrow$ AA) or R745A/Y746A (RY $\rightarrow$ AA) Myc-Nbs1, as indicated, were immunoprecipitated and immunoblotted.

complemented with vector or Nbs1 Δ C (Fig. 3b), as has been described for A-T cells²³. These findings suggest that an interaction between the Nbs1 C terminus and ATM is not needed for ATM activation *per se*; however, we cannot exclude a residual interaction mediated by the hypomorphic NBS1 allele expressed at low levels in NBS cells¹⁶. Nevertheless, these data clearly establish that the Nbs1 C-terminal ATM interaction motif is instrumental for effective phosphorylation of ATM targets at sites of DNA damage.

When damaged by ionizing radiation in the S or G2 phases of the cell cycle, mammalian cells transiently reduce DNA synthesis or block mitotic entry, respectively, in a manner partially dependent on Nbs1 (ref. 16). Notably, while cells complemented with wild-type Nbs1 efficiently downregulated DNA synthesis in response to ionizing radiation, cells complemented with vector or Nbs1 Δ C did this only weakly (Fig. 3c). Similarly, although the mitotic index rapidly dropped after irradiation of cells complemented with wild-type Nbs1, this response was severely impaired in vector- or Nbs1 Δ C-complemented cells (Fig. 3d). These data therefore indicate that the Nbs1–ATM interaction is required for the ionizing radiation-induced intra-S phase and G2/M checkpoints.

The Ku80 PIKK interaction motif is needed for DNA-PK function

The Ku heterodimer recruits DNA-PKcs to DNA ends to form the active DNA-PK complex^{11,12}, and the Ku80 C terminus has been implicated in this^{24,25}. Strikingly, in species that contain DNA-PKcs, Ku80 proteins contain a highly conserved C-terminal motif with homology to the C termini of both Nbs1 and ATRIP (Fig. 4a). Deletion or mutation of this Ku80 motif disrupted its interaction with DNA-PKcs (Fig. 4b). To study the function of this motif, we isolated DNA-binding protein complexes from complemented

Ku80-deficient Xrs-6Y cells (Fig. 4c, d). Although deletion of the last 14 residues from the Ku80 C terminus (Δ C) did not affect the formation or DNA-binding capabilities of the Ku heterodimer (Fig. 4c, d), DNA-PKcs only associated efficiently with DNA in extracts from cells expressing wild-type Ku80 (Fig. 4d). These results demonstrate that the conserved C-terminal motif of Ku80 is required for the efficient recruitment of DNA-PKcs to DNA ends *in vitro*.

When recruited to DNA ends and subsequently activated, DNA-PKcs phosphorylates itself on Thr2609 (refs 26, 27). Notably, autophosphorylated DNA-PKcs was readily detected in irradiated, wild-type-complemented Xrs-6Y cells but not in irradiated, vector- or Ku80 Δ C-complemented Xrs-6Y cells (Fig. 4e), revealing that the Ku80 C terminus is required for DNA-PKcs activation *in vivo*. Activation of DNA-PK is required for the efficient repair of radiation-induced DSBs, which can be measured by quantitative analysis of γ H2AX foci after irradiation²⁸. Such an analysis revealed that although wild-type-complemented Xrs-6Y cells displayed efficient repair of DSBs, repair was impaired in both vector- and Ku80 Δ C-complemented cells (Fig. 4f). Indeed, the repair defect of Ku80 Δ C-complemented cells was similar to that of DNA-PKcs-deficient V3 cells, but less severe than that of vector-complemented cells (Fig. 4f). Similarly, whereas Xrs-6Y cells expressing the Ku80 Δ C mutant were more radiation resistant than vector-complemented cells, they displayed a radiosensitivity similar to DNA-PKcs-deficient V3 cells but were more radiosensitive than Xrs-6Y cells expressing wild-type Ku80 (Fig. 4g). Together, these data confirm that recruitment of DNA-PKcs to DSBs is instrumental to its function, and demonstrate that this event requires the conserved motif at the extreme C terminus of Ku80.

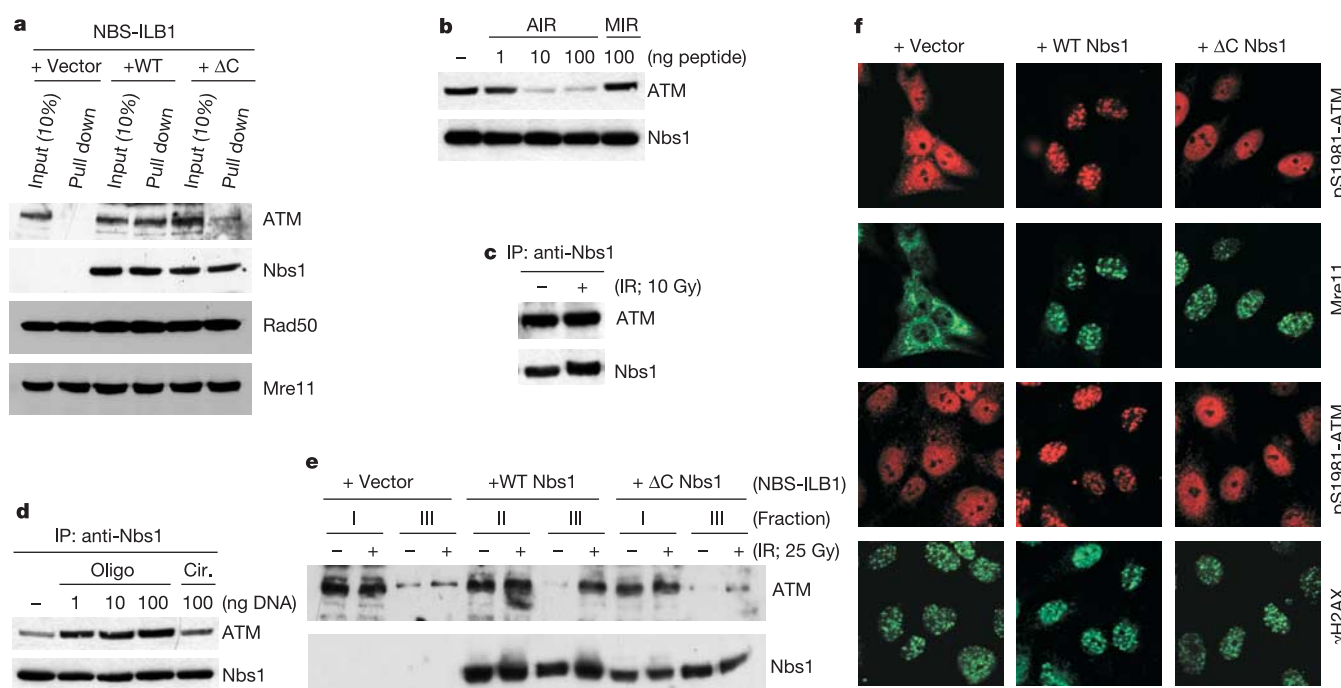


Figure 2 ATM association with sites of DNA damage requires its interaction with Nbs1. **a**, dsDNA bound proteins were isolated from extracts from complemented NBS-ILB1 cells and immunoblotted, as indicated. **b**, Nuclear extracts were pre-incubated with indicated amounts of peptide corresponding to the ATM or Mre11 interaction regions (AIR and MIR, respectively) of human Nbs1, and DNA bound proteins were isolated and immunoblotted as indicated. **c**, Extracts from untreated (–) or ionizing radiation-treated (10 Gy, 1 h) cells (+) were immunoprecipitated and immunoblotted as indicated. **d**, Nuclear extracts were

incubated with indicated amounts of a 30-mer dsDNA oligonucleotide (Oligo) or circular plasmid DNA (Cir.), then immunoprecipitated and immunoblotted as indicated. **e**, Complemented NBS-ILB1 cells were untreated (–) or irradiated (+) with 25 Gy of ionizing radiation. One hour later, cells were biochemically fractionated and extracts immunoblotted as indicated. Fractions I and III represent soluble and chromatin fractions, respectively. **f**, Complemented NBS-ILB1 cells were irradiated with 6 Gy of ionizing radiation. One hour after treatment, cells were fixed and immunostained as indicated.

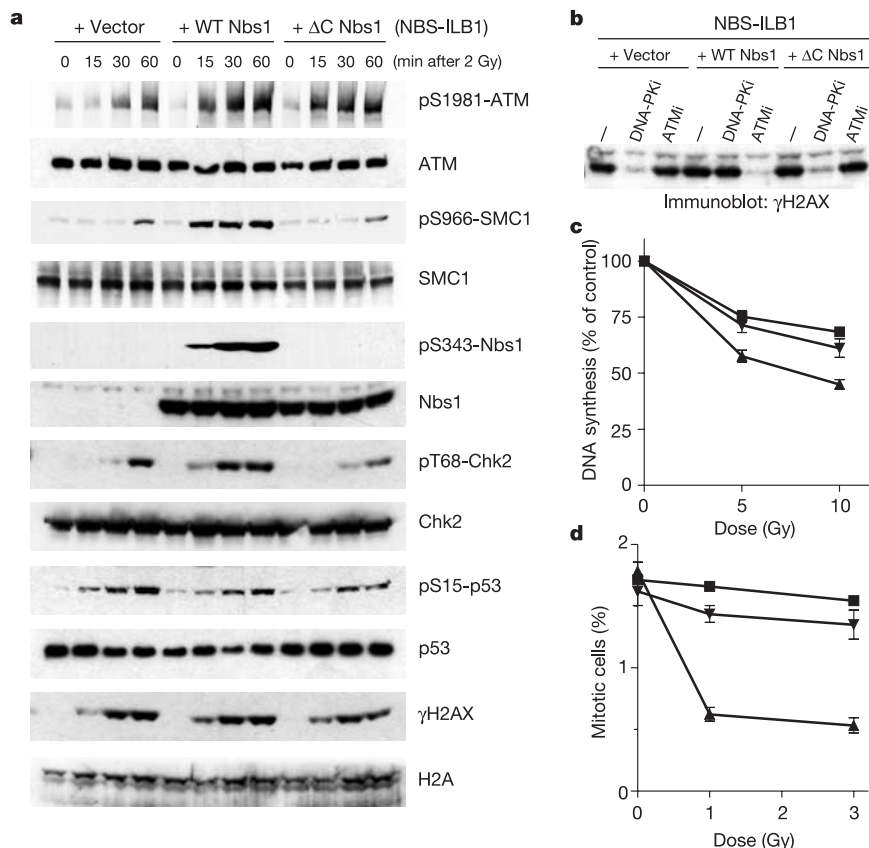


Figure 3 A functional Nbs1-ATM interaction is required for ATM phosphorylation events and checkpoint functions. **a**, Complemented NBS-ILB1 cells were treated with 2 Gy of ionizing radiation, extracts were prepared at the indicated time points after irradiation and immunoblotted as shown. **b**, Complemented NBS-ILB1 cells were pre-incubated with DMSO (–) or specific inhibitors of ATM (ATMi) or DNA-PKcs (DNA-PKi), irradiated with

2 Gy of ionizing radiation, and 1 h later, extracts were prepared and immunoblotted as indicated. **c**, **d**, The rate of DNA synthesis (**c**) and mitotic index (**d**) were determined 1 h after irradiation of vector- (filled squares), wild-type- (filled triangles) and Δ C (inverted triangles)-complemented NBS-ILB1 cells with the indicated doses of ionizing radiation. Error bars represent standard deviation.

a Human Nbs1	734	AKEE-SL	ADDLFRYN	746	(of 754)
Human ATRIP	767	DCEE-AA	DDLCAAE	780	(of 791)
Human Ku80	718	VFEEGGDV	DDLDDMI	732	(of 732)
Hamster Ku80	717	GPEEGGDV	DDLDDMI	731	(of 731)
Xenopus Ku80	712	MMEDEGDV	DDLDDMM	726	(of 726)

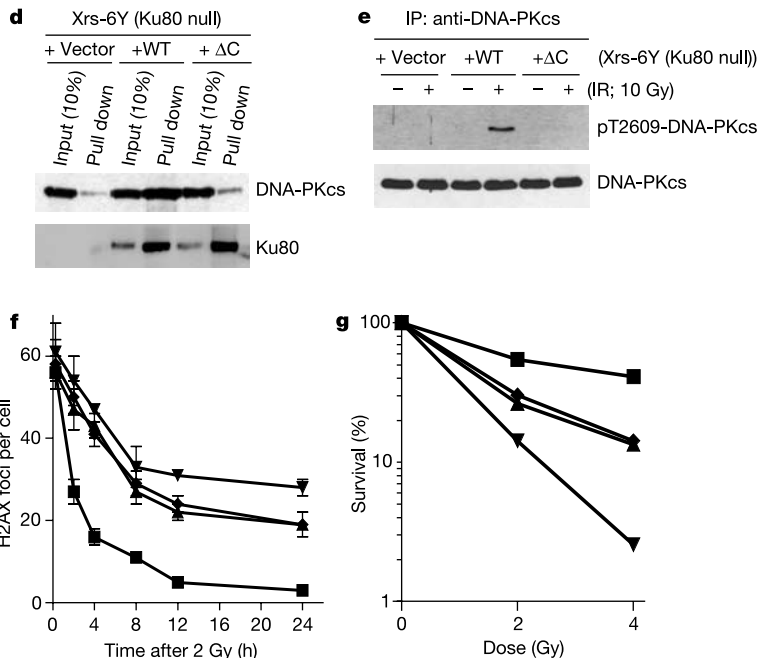
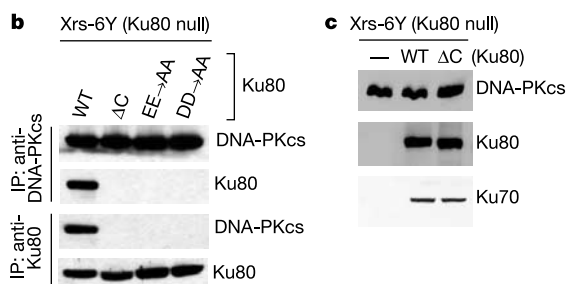


Figure 4 The PIKK interaction motif of Ku80 is required for DNA-PKcs activation. **a**, Alignment of the C termini of Nbs1, ATRIP and Ku80 from the indicated species. **b**, Extracts from Xrs-6Y cells expressing wild type, Δ C (amino acids 1–718), E720A/E721A (EE $\rightarrow$ AA) or D726A/D727A (DD $\rightarrow$ AA) Ku80 were immunoprecipitated and immunoblotted as indicated. **c**, Extracts from complemented Ku80-deficient Xrs-6Y cells were immunoblotted as indicated. Note that Ku70 stability depends on its interaction with Ku80. **d**, dsDNA bound proteins were isolated from extracts from complemented Xrs-6Y

cells and immunoblotted as indicated. **e**, Extracts from complemented Xrs-6Y cells were prepared 1 h after mock (–) or 10 Gy irradiation (+), then immunoprecipitated and immunoblotted as indicated. **f**, **g**, Temporal analysis of γ H2AX foci (**f**) and survival (**g**) of vector- (filled inverted triangles), wild-type-Ku80 (filled squares) and Ku80 Δ C (filled triangles)-complemented Xrs-6Y cells, and DNA-PKcs-deficient V3 cells (filled diamonds) exposed to the indicated doses of ionizing radiation. Error bars represent the standard deviation.

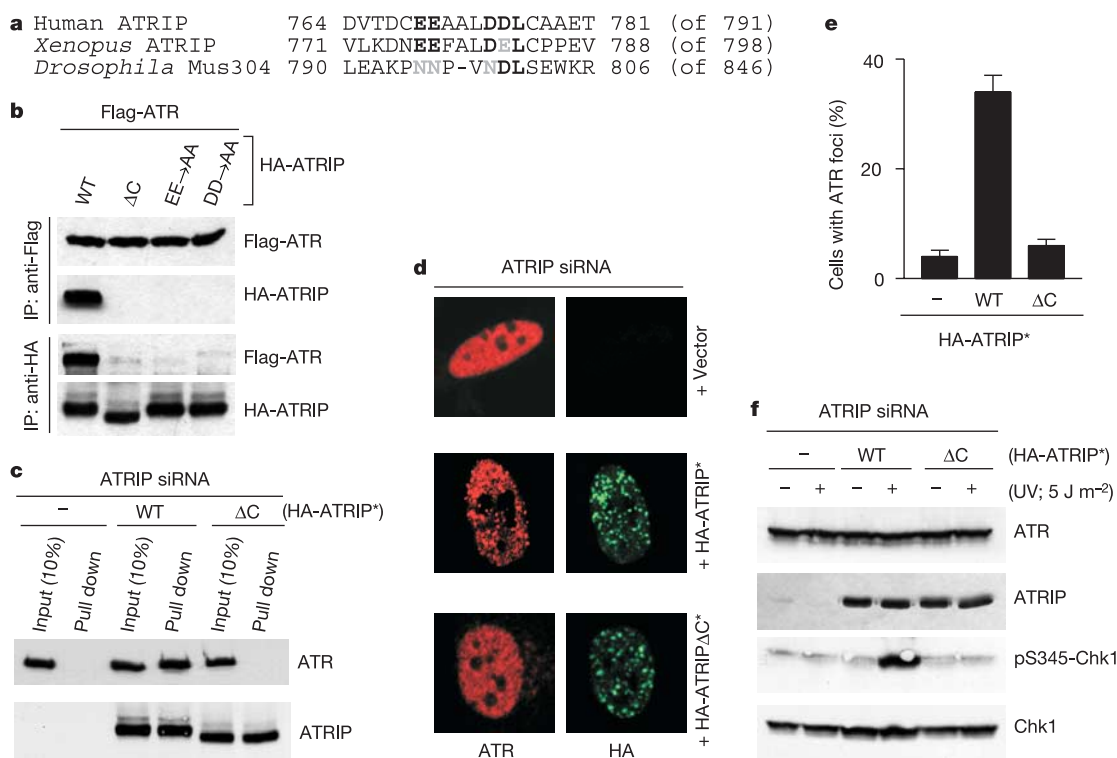


Figure 5 ATR-mediated signalling depends on the C-terminal PIKK interaction motif of ATRIP. **a**, Alignment of ATRIP C termini from the indicated species. **b**, Extracts from cells expressing Flag-ATR and wild-type, Δ C (amino acids 1–768), E769A/E770A (EE $\rightarrow$ AA) or D774A/D775A (DD $\rightarrow$ AA) HA-ATRIP were immunoblotted and immunoprecipitated as indicated. **c**, ssDNA bound proteins were isolated from 293T cells sequentially transfected (see Methods) with ATRIP siRNA and/or wild-type or Δ C HA-ATRIP*, as indicated, and

immunoblotted as indicated. **d**, **e**, HeLa cells sequentially transfected with ATRIP siRNA and/or wild-type or Δ C HA-ATRIP*, as indicated, were ultraviolet-irradiated with 10 J m⁻², fixed 2 h later and immunostained. Error bars represent the standard deviation. **f**, 293T cells were sequentially transfected with ATRIP siRNA and/or wild-type or Δ C HA-ATRIP* as indicated. One hour after 5 J m⁻² of ultraviolet irradiation, extracts were prepared and immunoblotted as indicated.

ATR function requires the PIKK interaction motif of ATRIP

To determine whether the conserved ATRIP C terminus (Fig. 5a) contributes to the ATR–ATRIP interaction, we deleted the 32 C-terminal residues of human ATRIP (Δ C). Notably, whereas wild-type ATRIP associated with ATR, ATRIP Δ C or ATRIP derivatives mutated in the C-terminal motif did not (Fig. 5b). To address whether loss of the ATR–ATRIP interaction would impair ATR function in human cells, we designed a short interfering RNA (siRNA) that targets ATRIP, and constructed an siRNA-resistant version of ATRIP (ATRIP*) by introducing silent mutations in the region of ATRIP targeted by the siRNA (Supplementary Fig. S3). As suggested by *in vitro* studies^{9,10}, ablation of ATRIP resulted in impaired binding of ATR to ssDNA (Fig. 5c). Expression of full-length ATRIP* in ATRIP siRNA-transfected cells restored ssDNA binding by ATR, but expression of ATRIP Δ C* did not (Fig. 5c). By contrast, binding of ATRIP to ssDNA was unaffected by the C-terminal deletion (Fig. 5c). Thus, a functional ATRIP–ATR interaction is required for recruitment of ATR to ssDNA. Consistent with these data, ATRIP knockdown severely impaired ultraviolet-light-induced focus formation by ATR, and ATR focus formation was restored by expression of full-length ATRIP* but not by ATRIP Δ C* (Fig. 5d, e).

ATR and its orthologues need to associate with DNA lesions to efficiently phosphorylate their target proteins^{10,29,30}. Hence, loss of ATRIP protein compromised ATR-mediated phosphorylation of Chk1 on Ser 345 after exposure to ultraviolet light (Fig. 5f). Importantly, ultraviolet-induced Chk1 phosphorylation by ATR was completely restored by expression of full-length ATRIP* but not by expression of ATRIP Δ C* (Fig. 5f). Collectively, these data demonstrate that disruption of the ATR–ATRIP complex by

deletion of the ATRIP C terminus prevents recruitment of ATR to sites of DNA damage and leads to defects in ATR-mediated phosphorylation events.

Discussion

We have shown that ATM, DNA-PKcs and ATR are recruited to sites of DNA damage through analogous mechanisms involving conserved interaction motifs present in Nbs1, Ku80 and ATRIP, respectively. As DNA damage recruitment is key to the functions of these molecules, our findings have important mechanistic, biological and evolutionary implications for our understanding of DNA damage-initiated events.

First, we have established that the human MRN complex recruits ATM to sites of damage through a motif at the Nbs1 C terminus. Analogous motifs exist at the C termini of all Nbs1 orthologues, suggesting that this role has been conserved throughout eukaryotic evolution. Consistent with this idea, recruitment of *Xenopus* ATM to a DNA-bound signalling complex in *Xenopus* egg extracts was shown to require *Xenopus* Mre11 (ref. 31), and a functional Xrs2–Tel1 interaction is needed for Tel1 association with DSBs *in vivo*³². The mode of ATM recruitment to DNA ends by Nbs1 resembles the Ku- and ATRIP-mediated association of DNA-PKcs and ATR, respectively, with DNA lesions. As this suggests an underlying mechanistic basis for these similarities, we have identified conserved C-terminal sequence motifs in Ku80 and ATRIP that resemble the ATM interaction motif of Nbs1. Indeed, deletion of the Ku80 motif impairs the ability of Ku to recruit DNA-PKcs to DNA ends and thereby prevents DNA-PK activation and efficient DSB repair *in vivo*. Correspondingly, deletion or mutation of the analogous motif in ATRIP disrupts its interaction with ATR,

impairs ATR recruitment to ssDNA and ultraviolet-induced DNA lesions, and prevents ATR phosphorylating its key downstream target, Chk1. Previous studies in yeast and *Drosophila* have demonstrated that subtle C-terminal deletions of ATRIP orthologues not only resulted in ATRIP null phenotypes but also mirrored ATR deficiency^{33–35}. In all of these cases, the ATRIP C-terminal PIKK interaction motifs were deleted, thus providing strong support for functional conservation of this ATR–ATRIP interaction mode throughout evolution.

Although our data indicate related mechanisms for DNA damage recruitment by ATM, ATR and DNA-PKcs, there are apparent differences in the way that these interactions take place. Assembly of the DNA-PKcs–Ku complex requires linear DNA, suggesting that Ku binds to DNA ends first and then recruits DNA-PKcs^{11,12}. By contrast, ATR and ATRIP are thought to arrive at DNA lesions as a pre-assembled complex^{9,10}. Because interaction between MRN and ATM is apparently DNA independent and unaffected by ionizing radiation (Fig. 2c; see also ref. 36), MRN and ATM might also arrive at sites of DNA damage as a pre-assembled complex. However, MRN association with DSBs does not require ATM, suggesting that MRN and ATM are sequentially recruited to sites of DNA damage. Consistent with this idea, we have found that DNA binding by MRN enhances its affinity for ATM (Fig. 2d). Recent work has also shown that ATM function is modulated by a two-step mechanism: the first involves autophosphorylation of Ser 1981, which dissociates an inactive ATM dimer into active monomers; the second involves the recruitment of this active ATM to sites of DNA damage^{15,17}. Our data suggest that although the Nbs1 C terminus is not required for ATM autophosphorylation in human cells, it is crucial for recruiting activated ATM to sites of DSBs in order to facilitate the phosphorylation of ATM targets at such sites. Thus, DSB-induced ATM autophosphorylation is likely to require features of the MRN complex that operate independently of the Nbs1 C terminus, such as its DNA binding and nuclease activities.

The presence of the PIKK interaction motifs in the extreme C termini of Nbs1, Ku80 and ATRIP might make them easily accessible. Notably, we have found that although the C-terminal motifs of Nbs1, Ku80 and ATRIP display some selectivity towards their respective PIKK within the context of short peptides, full PIKK discrimination is only imparted when the motifs are presented in the context of the full-length PIKK interaction partner (Supplementary Fig. S4). Therefore, the C-terminal PIKK interaction motif probably serves as a kinase-docking module that is required for the subsequent establishment of more stable associations involving multiple regions of interaction. Interestingly, we have mapped the Nbs1 interaction region of ATM to a few clusters of HEAT repeats (Supplementary Fig. S5). Although the HEAT-repeat-containing regions are not highly conserved between the PIKKs, they have recently been suggested to adopt similar structures^{37,38}. As the PIKK interaction motifs of Nbs1, Ku80 and ATRIP display sequence similarity with each other, it is tempting to speculate that ATM, DNA-PKcs and ATR might contain reciprocal interaction regions of significant structural conservation.

We have identified conserved and functionally related motifs in Nbs1, Ku80 and ATRIP required for their interaction with ATM, DNA-PKcs and ATR, respectively. Deletion of these motifs impairs recruitment of the PIKKs to DNA lesions and severely compromises PIKK-dependent signalling events. Our findings not only directly demonstrate that PIKK recruitment is essential for PIKK-mediated checkpoint and repair functions, but also reveal that the mechanisms of PIKK recruitment display more general features than has hitherto been anticipated. □

Methods

Cell culture

SV40-transformed NBS-ILB1 fibroblasts and Ku80-deficient Xrs-6Y hamster cells stably expressing human DNA-PKcs (gift from P. Jeggo³⁹), derivatives thereof, HeLa and 293T

cells were cultured in DMEM containing 10% serum. Complementation of NBS-ILB1 and Xrs-6Y cells was done by transfection of pBabe-puro plasmids encoding wild-type or Δ C versions of Nbs1 and Ku80, followed by selection with puromycin. Transfection of plasmids and siRNA was done using Lipofectamine 2000 and Oligofectamine (both from Invitrogen), respectively. Sequential transfection of 293T or HeLa cells was performed by first transfecting cells with siRNAs and, after a 16-h incubation, transfecting with siRNA-resistant HA-ATRIP* vectors (see below). The luciferase- (CGUACGCGAAUACUUCG AdTdT) and ATRIP-targeting (AGAGAAACUGUCCAAUAdTdT) siRNA duplexes were from Dharmacon. Highly specific small-molecule inhibitors of ATM (KU55933; ref. 21) and DNA-PKcs (NU7026; ref. 22) were used at 10 μ M (gift from G. Smith). X-ray and ultraviolet irradiation of cells, clonogenic survival, radio-resistant DNA synthesis and G2/M assays were done as previously described^{40–42}. Cell extracts were prepared in SDS sample buffer or buffer A (50 mM Tris-HCl, pH 8.0, 120 mM NaCl, 0.5% NP-40, 1 mM EDTA, 1 mM NaF, 1 mM β -glycerophosphate, 0.1 mM Na₃VO₄, 1 mM dithiothreitol (DTT), 10 μ g ml⁻¹ leupeptin, 1 mM AEBF). ATM chromatin retention assays were performed essentially as described¹³. Sequential extraction with NP-40 yielded three fractions (I, II and III).

Plasmids

Expression vectors for Myc-Nbs1, Flag-ATM and Flag-ATR have been described^{9,41}. The haemagglutinin-tagged (HA) ATRIP expression vector was a gift from P. M. Reaper. The pBabe-puro Nbs1 Δ C, wild-type Ku80 and Ku80 Δ C, and HA-ATRIP Δ C vectors were generated by polymerase chain reaction (PCR) cloning. To generate RNA interference-resistant ATRIP, the region targeted by the ATRIP siRNA (see above) was changed to AGAGAGACgTcCgAATA. Mutagenesis of Nbs1, Ku80 and ATRIP was done using the QuikChange Kit (Stratagene).

Antibodies and immunofluorescence

Antibodies used in this study were from Novus Biologicals (ATM, Mre11 and Rad50), Upstate (γ H2AX), Rockland Immunochemicals (pS1981-ATM), Bethyl Laboratories (pS966-SMC1 and SMC1), Roche (HA and Myc), Neomarkers (DNA-PKcs and Ku80), Cell Signalling Technology (pT68-Chk2, pS345-Chk1 and pS15-p53), Sigma (Flag M2) and Santa Cruz Biotechnology (p53, Chk1, Chk2, H2A and ATR). Rabbit antiserum to ATR used in immunofluorescence was a gift from R. Abraham⁴³. Rabbit antisera to DNA-PKcs, Ku80 and Ku70, and sheep antiserum against Mre11, have been described elsewhere^{24,40}. Rabbit antiserum to ATRIP was raised against a recombinant fragment of human ATRIP. For immunofluorescent staining, cells were fixed in 4% paraformaldehyde, permeabilized in 0.2% Triton X-100, and immunostained with antibodies as indicated.

Peptide and DNA pull downs

Biotinylated peptides were synthesized at the Department of Biochemistry, University of Bristol, UK. The sequences were BIOTIN-SGSINDDYGLKFNFKKVTYPGAG (Nbs1 675–697), BIOTIN-SGSKEESLADDFRYNPYLKRRR (Nbs1 734–754) and BIOTIN-SGSGDGDGDDDDGPKFTFKRRKG (Xrs2 835–854). Peptides were coupled to streptavidin-coated Dynabeads M-280 (Dyna), incubated with HeLa nuclear extracts, washed with Tris-buffered saline, and re-suspended in SDS sample buffer before immunoblotting. For DNA pull downs, a 250-bp fragment, generated by PCR using a biotinylated primer, or a 75-mer biotinylated oligonucleotide were used¹⁹. The DNA was coupled to streptavidin-coated Dynabeads M-280 (Dyna), incubated with cell extracts prepared in buffer B (50 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM MgCl₂, 5 mM EGTA, 1 mM DTT, 0.5% Triton X-100, 10% glycerol, 1 mM NaF, 1 mM β -glycerophosphate, 0.1 mM Na₃VO₄, 1 mM DTT, 10 μ g ml⁻¹ leupeptin, 1 mM AEBF), washed in Tris-buffered saline and subjected to immunoblotting.

Received 7 December 2004; accepted 14 February 2005; doi:10.1038/nature03442.

Published online 2 March 2005.

1. Zhou, B.-B. S. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408**, 433–439 (2000).
2. Rouse, J. & Jackson, S. P. Interfaces between the detection, signaling, and repair of DNA damage. *Science* **297**, 547–551 (2002).
3. Melo, J. & Toczyski, D. A unified view of the DNA-damage checkpoint. *Curr. Opin. Cell Biol.* **14**, 237–245 (2002).
4. Shiloh, Y. ATM and related protein kinases: safeguarding genome integrity. *Nature Rev. Cancer* **3**, 155–168 (2003).
5. Meek, K., Gupta, S., Ramsden, D. A. & Lees-Miller, S. P. The DNA-dependent protein kinase: the director at the end. *Immunol. Rev.* **200**, 132–141 (2004).
6. Hammarsten, O., DeFazio, L. G. & Chu, G. Activation of DNA-dependent protein kinase by single-stranded DNA ends. *J. Biol. Chem.* **275**, 1541–1550 (2000).
7. Smith, G. C. M. *et al.* Purification and DNA binding properties of the ataxia-telangiectasia gene product ATM. *Proc. Natl Acad. Sci. USA* **96**, 11134–11139 (1999).
8. Ünsal-Kaçmaz, K., Makhov, A. M., Griffith, J. D. & Sancar, A. Preferential binding of ATR protein to UV-damaged DNA. *Proc. Natl Acad. Sci. USA* **99**, 6673–6678 (2002).
9. Ünsal-Kaçmaz, K. & Sancar, A. Quaternary structure of ATR and effects of ATRIP and replication protein A on its DNA binding and kinase activities. *Mol. Cell. Biol.* **24**, 1292–1300 (2004).
10. Zou, L. & Elledge, S. J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* **300**, 1542–1548 (2003).
11. Gottlieb, T. M. & Jackson, S. P. The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* **72**, 131–142 (1993).
12. Dvir, A., Peterson, S. R., Knuth, M. W., Lu, H. & Dynan, W. S. Ku autoantigen is the regulatory component of a template-associated protein kinase that phosphorylates RNA polymerase II. *Proc. Natl Acad. Sci. USA* **89**, 11920–11924 (1992).
13. Andegeko, Y. *et al.* Nuclear retention of ATM at sites of DNA double strand breaks. *J. Biol. Chem.* **276**, 38224–38230 (2001).

14. Carson, C. T. *et al.* The Mre11 complex is required for ATM activation and the G2/M checkpoint. *EMBO J.* **22**, 6610–6620 (2003).
15. Kitagawa, R., Bakkenist, C. J., McKinnon, P. J. & Kastan, M. B. Phosphorylation of SMC1 is a critical downstream event in the ATM-NBS1-BRCA1 pathway. *Genes Dev.* **18**, 1423–1438 (2004).
16. D'Amours, D. & Jackson, S. P. The Mre11 complex: at the crossroads of DNA repair and checkpoint signalling. *Nature Rev. Mol. Cell Biol.* **3**, 317–327 (2002).
17. Bakkenist, C. J. & Kastan, M. B. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003).
18. Lukas, C., Falck, J., Bartkova, J., Bartek, J. & Lukas, J. Distinct spatiotemporal dynamics of mammalian checkpoint regulators induced by DNA damage. *Nature Cell Biol.* **5**, 255–260 (2003).
19. Uziel, T. *et al.* Requirement of the MRN complex for ATM activation by DNA damage. *EMBO J.* **22**, 5612–5621 (2003).
20. Horejsi, Z. *et al.* Distinct functional domains of Nbs1 modulate the timing and magnitude of ATM activation after low doses of ionizing radiation. *Oncogene* **23**, 3122–3127 (2004).
21. Hickson, I. *et al.* Identification and characterization of a novel and specific inhibitor of the ataxia-telangiectasia mutated kinase ATM. *Cancer Res.* **64**, 9152–9159 (2004).
22. Veuger, S. J., Curtin, N. J., Smith, G. C. & Durkacz, B. W. Effects of novel inhibitors of poly(ADP-ribose) polymerase-1 and the DNA-dependent protein kinase on enzyme activities and DNA repair. *Oncogene* **23**, 7322–7329 (2004).
23. Stiff, T. *et al.* ATM and DNA-PK function redundantly to phosphorylate H2AX after exposure to ionizing radiation. *Cancer Res.* **64**, 2390–2396 (2004).
24. Gell, D. & Jackson, S. P. Mapping of protein–protein interactions within the DNA-dependent protein kinase complex. *Nucleic Acids Res.* **27**, 3494–3502 (1999).
25. Singleton, B. K., Taccioli, G. E., Rottinghaus, S. & Jeggo, P. A. Interaction of the C-terminus of Ku80 with the DNA-dependent protein kinase catalytic subunit. *Mol. Cell Biol.* **19**, 3267–3277 (1999).
26. Chan, D. W. *et al.* Autophosphorylation of the DNA-dependent protein kinase catalytic subunit is required for rejoining of DNA double-strand breaks. *Genes Dev.* **16**, 2333–2338 (2002).
27. Douglas, P. *et al.* Identification of *in vitro* and *in vivo* phosphorylation sites in the catalytic subunit of the DNA-dependent protein kinase. *Biochem. J.* **368**, 243–251 (2002).
28. Rothkamm, K. & Lobrich, M. Evidence for a lack of DNA double-strand break repair in human cells exposed to very low x-ray doses. *Proc. Natl Acad. Sci. USA* **100**, 5057–5062 (2003).
29. Zou, L., Cortez, D. & Elledge, S. J. Regulation of ATR substrate selection by Rad17-dependent loading of Rad9 complexes onto chromatin. *Genes Dev.* **16**, 198–208 (2002).
30. Rouse, J. & Jackson, S. P. Lcd1p recruits Mec1p to DNA lesions *in vitro* and *in vivo*. *Mol. Cell* **9**, 857–869 (2002).
31. Costanzo, V., Paull, T., Gottesman, M. & Gautier, J. Mre11 assembles linear DNA fragments into DNA damage signaling complexes. *PLoS Biol.* **2**, E110 (2004).
32. Nakada, D., Matsumoto, K. & Sugimoto, K. ATM-related Tel1 associates with double-strand breaks through an Xrs2-dependent mechanism. *Genes Dev.* **17**, 1957–1962 (2003).
33. Brodsky, M. H., Sekelsky, J. J., Tsang, G., Hawley, R. S. & Rubin, G. M. *mus304* encodes a novel DNA damage checkpoint protein required during *Drosophila* development. *Genes Dev.* **14**, 666–678 (2000).
34. Wakayama, T., Kondo, T., Ando, S., Matsumoto, K. & Sugimoto, K. Pie1, a protein interacting with Mec1, controls cell growth and checkpoint responses in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **21**, 755–764 (2001).
35. Wolkow, T. D. & Enoch, T. Fission yeast Rad26 is a regulatory subunit of the Rad3 checkpoint kinase. *Mol. Biol. Cell* **13**, 480–492 (2002).
36. Lee, J. H. & Paull, T. T. Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex. *Science* **304**, 93–96 (2004).
37. Brewerton, S. C., Dore, A. S., Drake, A. C., Leuther, K. K. & Blundell, T. L. Structural analysis of DNA-PKcs: modelling of the repeat units and insights into the detailed molecular architecture. *J. Struct. Biol.* **145**, 295–306 (2004).
38. Perry, J. & Kleckner, N. The ATRs, ATMs, and TORs are giant HEAT repeat proteins. *Cell* **112**, 151–155 (2003).
39. Singleton, B. K., Torres-Arzayus, M. I., Rottinghaus, S. T., Taccioli, G. E. & Jeggo, P. A. The C terminus of Ku80 activates the DNA-dependent protein kinase catalytic subunit. *Mol. Cell Biol.* **19**, 3267–3277 (1999).
40. Goldberg, M. *et al.* MDC1 is required for the intra-S-phase DNA damage checkpoint. *Nature* **421**, 952–956 (2003).
41. Falck, J., Petrini, J. H. J., Williams, B. R., Lukas, J. & Bartek, J. The DNA damage-dependent intra-S phase checkpoint is regulated by parallel pathways. *Nature Genet.* **30**, 290–294 (2002).
42. Xu, B., Kim, S.-T. & Kastan, M. B. Involvement of Brca1 in S-phase and G₂-phase checkpoints after ionizing irradiation. *Mol. Cell Biol.* **21**, 3445–3450 (2001).
43. Tsao, C. C., Geisen, C. & Abraham, R. T. Interaction between human MCM7 and Rad17 proteins is required for replication checkpoint signaling. *EMBO J.* **23**, 4660–4669 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to K.-P. Hopfner and members of the Jackson laboratory for their suggestions, V. Smits for help with the G2/M checkpoint assay, and P. Jeggo, R. T. Abraham, G. Smith, D. J. Chen, H. H. Chun and R. A. Gatti for providing reagents. We thank M. Stucki and P. M. Reaper for critical reading of the manuscript. This study was supported by Cancer Research UK and by the Danish Cancer Society (to J.F.).

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com.

Correspondence and requests for materials should be addressed to S.P.J. (s.jackson@gurdon.cam.ac.uk).

Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA

Anirban Banerjee¹, Wei Yang¹, Martin Karplus^{1,3} & Gregory L. Verdine^{1,2}

¹Departments of Chemistry and Chemical Biology, and ²Molecular and Cellular Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

³L'Institut de Science et d'Ingénierie Supramoléculaires (ISIS), Université Louis Pasteur, 8 allée Gaspard Monge, 67000 Strasbourg, France

How DNA repair proteins distinguish between the rare sites of damage and the vast expanse of normal DNA is poorly understood. Recognizing the mutagenic lesion 8-oxoguanine (oxoG) represents an especially formidable challenge, because this oxidized nucleobase differs by only two atoms from its normal counterpart, guanine (G). Here we report the use of a covalent trapping strategy to capture a human oxoG repair protein, 8-oxoguanine DNA glycosylase I (hOGG1), in the act of interrogating normal DNA. The X-ray structure of the trapped complex features a target G nucleobase extruded from the DNA helix but denied insertion into the lesion recognition pocket of the enzyme. Free energy difference calculations show that both attractive and repulsive interactions have an important role in the preferential binding of oxoG compared with G to the active site. The structure reveals a remarkably effective gate-keeping strategy for lesion discrimination and suggests a mechanism for oxoG insertion into the hOGG1 active site.

OxoG, produced by the attack of intracellular oxidants on G residues in DNA^{1,2}, is misread as thymine (T) by the replicative DNA polymerization machinery. The resulting oxoG•A mispair, if allowed to persist, gives rise through one further round of replication to a G•C to T•A transversion mutation. It is the role of hOGG1 to intercept the oxoG•C lesions once formed and initiate correction by the base-excision DNA repair pathway (Fig. 1a) before they are encountered by the replication machinery. Only 50,000 molecules of hOGG1 protect the entire 6×10^9 base-pair nuclear genome of a diploid human cell³, hence the enzyme must have developed an efficient mechanism for distinguishing oxoG from the four nucleobases in normal DNA.

The structural basis for recognition and removal of oxoG by hOGG1 has been intensively studied^{4–7}. The structure of recognition-competent but repair-defective mutants of hOGG1 bound to oxoG-containing DNA revealed⁴ that the target oxoG nucleoside is completely extruded from the DNA duplex and inserted deeply into a lesion recognition pocket on the enzyme. Amino acid residues lining this pocket contact oxoG directly, providing a basis for specific recognition of the lesion (Fig. 1a, inset). Because hOGG1 also makes extensive contacts to the C residue left unpaired by oxoG extrusion ('estranged C'; see below), the problem of lesion discrimination by the hOGG1 catalytic machinery is reduced to understanding how the lesion recognition pocket distinguishes oxoG from G. These two nucleobases differ in chemical composition at only two positions: C8 (O versus H) and N7 (H versus lone electron pair) (Fig. 1a). The only obvious contribution to discrimination apparent from structural studies is the hydrogen bond made by the N7 H of oxoG to the main-chain carbonyl of Gly 42 (Fig. 1a), which would be lost with G. Is this one single interaction sufficient to enable discrimination of oxoG from G, despite a roughly million-fold concentration difference between the two? To address this question, we sought to compare the available structures of hOGG1–DNA complexes presenting an oxoG lesion to the active site with a structure presenting the normal base G, and to use the set of structures as a basis for free energy difference calculations to elucidate the nature of the important interactions. Obtaining the latter structure represented a formidable technical challenge, because of necessity, DNA-binding proteins generally fail to form homogeneous complexes with DNA lacking cognate structural

features such as lesions or promoter sequences. Consequently, few structures of proteins bound nonspecifically to DNA have been determined so far^{8–11}.

In previous work^{12,13}, we reported the development of a method by which to impose structural homogeneity on otherwise transient or inhomogeneous protein–DNA interaction systems. This method, disulphide trapping¹⁴, consists of implanting a disulphide crosslink into the interface between a protein and DNA, so as to restrict the ability of the two to dissociate from each other and hence form alternative complexes. Here we report the use of disulphide trapping to observe hOGG1 in the act of presenting G to its active site. Together with the accompanying computation studies, the present structures reveal that hOGG1 uses a refined mechanism for denying G access to the active site. The structure with the normal base bound to an alternative nucleobase-binding site suggests a possible late intermediate in the base-extrusion pathway.

Overall experimental strategy

The overall experimental strategy for these studies is outlined in Fig. 1b–d. Briefly, the structure of the previously reported hOGG1–oxoG DNA lesion recognition complex⁴ was inspected in order to identify prospective crosslinking sites. Of the several candidate crosslinking positions in the protein thus identified, position 149 appeared especially attractive, because it lies completely outside the enzyme active site and yet constitutes a point of intimate contact between the protein and DNA. Specifically, the side-chain carbonyl of Asn 149 hydrogen bonds to the exocyclic amine of the estranged cytosine, the base left unpaired by extrusion of oxoG from the DNA helix (Fig. 1c; see also Supplementary Fig. S1a). We reasoned that replacement of this hydrogen-bonding interaction with a covalent crosslink (Fig. 1c) might have minimal impact on the local conformation at the site of crosslinking, while ensuring extrusion of the complementary guanine from the DNA helix. However, the feasibility of crosslinking at this site in non-lesion-containing DNA was uncertain, because the crosslinking reaction would require sustained disruption of a normal C•G pair located amidst 14 neighbouring base pairs that could equally well be disrupted.

Crosslinking chemistry and structural validation

To test the present crosslinking strategy, we started with the

catalytically inactive K249Q variant of hOGG1 and introduced a further N149C mutation to give hOGG1-QC. We constructed a series of duplex oligonucleotides having the sequence shown in Fig. 1d, with a single oxoG lesion in the upper strand and thiol-bearing tether (activated as a mixed disulphide) attached to the estranged cytosine on the lower strand (as in Fig. 1c, $n = 1$). Upon incubation of these oligonucleotides with hOGG1-QC, significant amounts of a crosslinked species were formed (Supplementary Fig. S1b); crosslinking was dependent upon the N149C mutation and was lost upon incubation with 2-mercaptoethanol (data not shown).

Inspection of the structure of the crosslinked oxoG complex refined to 2.4 Å (Supplementary Table) revealed that the active site was identical within experimental error to that of the un-crosslinked oxoG recognition complex⁴, with the extra-helical oxoG nucleoside deeply inserted into the concave lesion recognition pocket (data not shown); only minor differences were observed elsewhere, and these were localized entirely to the immediate vicinity of the crosslink (Supplementary Fig. S1a, compare the top pair of structures). Crosslinking is associated with a slight retraction of the estranged cytosine from the protein surface, but the set of hydrogen-bonding contacts to Arg 154 and Arg 204 appears to be maintained nonetheless. Electron density for the thiol-bearing tether on DNA is weaker than that for most of the surrounding atoms in the structure, indicating conformational flexibility in the crosslink itself; observations of weak electron density for similar disulphide crosslinks have been made previously^{12,13,15}. These results establish the suitability of the Cys 149/estranged C crosslink for positional trapping of hOGG1–DNA complexes.

Crosslinking hOGG1 to normal DNA

The oxoG lesion in the oligonucleotide was replaced with G, and crosslinking was performed using a version of hOGG1 containing only the N149C mutation (N149C hOGG1). Although crosslinking is considerably slower with the G-containing duplex than with oxoG (Supplementary Fig. S1b), significant amounts of crosslinked product were obtained. Crosslinking was dependent upon the N149C mutation (data not shown) and tether length (Supplementary Fig. S1c), demonstrating selectivity in the crosslinking reaction.

The global structure of hOGG1 in complex with a non-lesion-containing DNA duplex (G complex) refined to 2.50 Å resolution (Supplementary Table) bears marked overall similarity to that of the oxoG complex (Fig. 2), the most distinctive feature being the presentation of an extra-helical nucleoside to the hOGG1 active site by a drastically bent duplex ($\sim 80^\circ$ compared with $\sim 70^\circ$, respectively). The local structure at the site of crosslinking is largely unaffected by changing the extra-helical oxoG to G (Supplementary Fig. S1a). Whereas the oxoG nucleobase inserts itself deeply into the lesion recognition pocket on the enzyme (Figs 2, left, and 3a), G is rejected by the lesion recognition pocket and instead lies against the protein surface at an exo-site some 5 Å outside the pocket (Figs 2, right panel, and 3b). The G base does interact with two active-site residues, Phe 319 and His 270, but the contacts are completely different from those made with oxoG. In the oxoG complex, His 270 does not contact the oxoG nucleobase, but instead hydrogen bonds to its 5' phosphate (Fig. 3a); in the G complex, this phosphate contact is disengaged, and His 270 interacts with the π -face of the G nucleobase (Fig. 3b). The high *B*-factor for the His 270 side chain in the G complex suggests that its interaction with G is weak. These

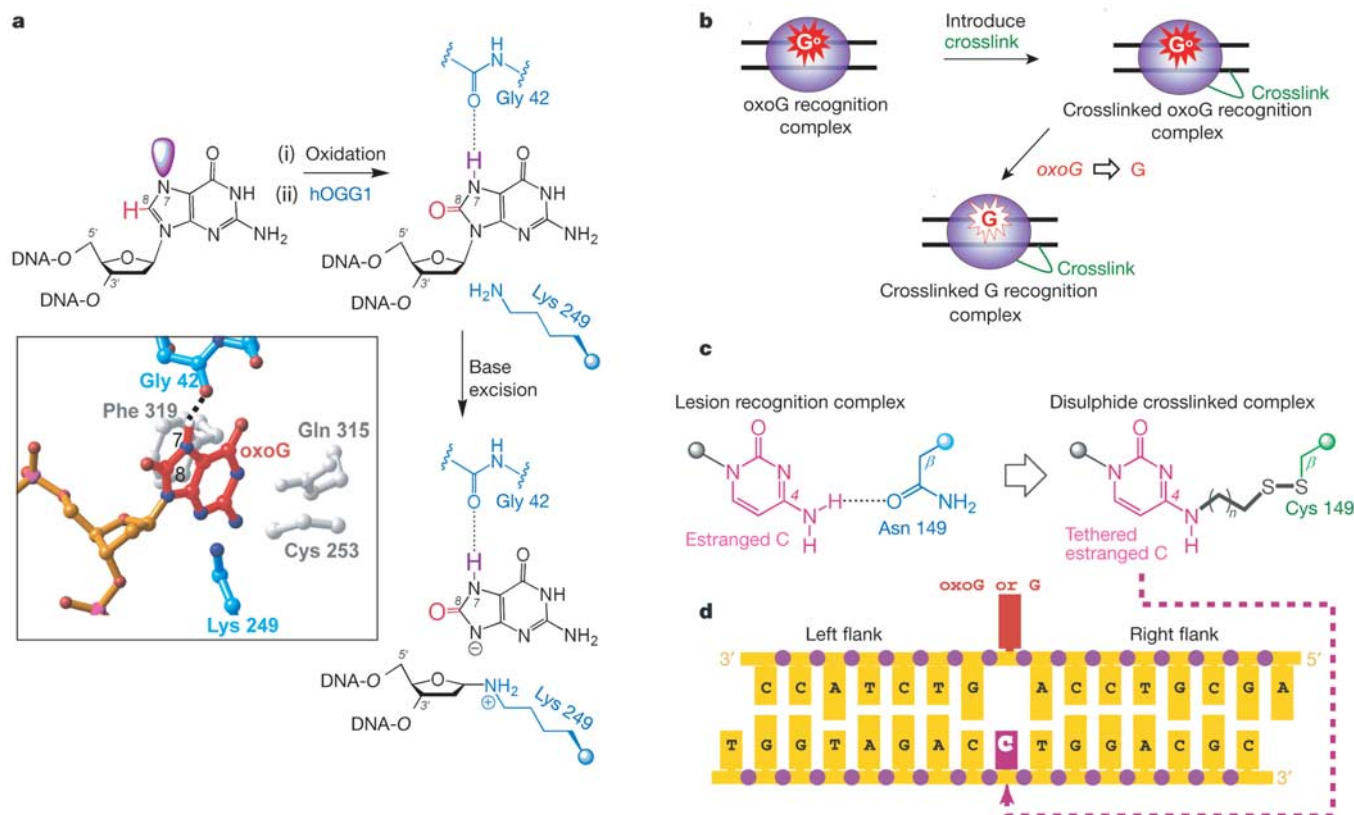


Figure 1 Generation of 8-oxoguanine (oxoG), its recognition by human 8-oxoguanine DNA glycosylase (hOGG1) and overview of the structure-based trapping strategy used here to obtain a complex of hOGG1 bound to undamaged DNA. **a**, Structural differences between G and oxoG. Inset: close-up view of the lesion recognition pocket of hOGG1 showing residues involved in recognition of oxoG and catalysis, highlighting the direct contact

between N7 H and Gly 42, and the catalytic nucleophile Lys 249. **b**, Schematic overview of the crosslinking and validation strategy. **c**, Details of the trapping chemistry. Attachment of a tether at the N4 position of cytosine is known to preserve Watson–Crick pairing in protein-unbound DNA, with the tether protruding into the major groove³¹. **d**, Sequence of the DNA duplex used in this work.

results establish that hOGG1 is able to read out the subtle structural distinctions between the oxoG lesion versus its normal counterpart G, allowing the lesion admittance to the active-site pocket while rejecting its normal counterpart.

Despite having a fully functional catalytic apparatus, the G complex undergoes no detectable base excision over the course of 10 days (data not shown). A version of the oxoG complex having Lys 249 intact (that is, constructed with N149C hOGG1) does undergo base excision, indicating that crosslinking per se does not abrogate catalysis. We thus conclude that hOGG1 is unable to cleave G from DNA at a detectable rate, even with the benefit of having G forcibly presented to the enzyme active site.

Energetic basis for discrimination of oxoG from G

The structure of the G complex presented here demonstrates the existence of an alternative nucleobase-binding site outside the lesion recognition pocket (exo-site). Because there are two sites on the enzyme that could, in principle, bind oxoG and G, the quantity of thermodynamic interest is the discrimination factor for binding oxoG versus G in the lesion recognition pocket relative to the differential binding of the two moieties in the exo-site. Free energy simulations (see Methods) were performed to determine this factor and dissect it into the important contributions. Calculations of the free energy penalties of inserting G versus oxoG in the two sites (ΔA_1 and ΔA_2 in Fig. 4a) is difficult to do directly by computational methods because of the large conformational change involved in the process. This problem was overcome by constructing a thermodynamic cycle (Fig. 4a), with which the overall discrimination free energy $\Delta\Delta A = \Delta A_1 - \Delta A_2$ can be obtained by comparing the free energy costs of bringing about an 'alchemical' transformation of oxoG to G in the active site (ΔA_3) and the exo-site (ΔA_4); such alchemical transformations can be simulated accurately by well-defined methods^{16,17}. The calculated value for $\Delta\Delta A$ was found to be $-6.8 \text{ kcal mol}^{-1}$, which corresponds to a roughly 10^5 -fold preference to insert oxoG versus G into the active site, starting with either nucleobase at the exo-site observed in the G complex. To determine whether the main contribution to the discrimination factor arises from the relative binding of oxoG versus G to the active site, the exo-site, or both, additional alchemical simulations were performed (see Supplementary Information). As

shown by a free energy simulation that uses aqueous solution as the reference, the active site favours oxoG relative to G by $8.6 \text{ kcal mol}^{-1}$. As the calculated discrimination factor is only $-6.8 \text{ kcal mol}^{-1}$, the exo-site also favours oxoG by the difference, which is equal to $1.8 \text{ kcal mol}^{-1}$.

The simulations for the relative binding of oxoG versus G to the hOGG1 active site have Lys 249 in the fully protonated state ($-\text{NH}_3^+$) and Cys 253 in the deprotonated form (S^-), with the two forming a salt bridge (Fig. 4b). Our calculations show that this configuration is more stable than the alternative neutral pair Lys 249 (NH_2)/Cys 253 (SH) or the pair having both Lys 249 and Cys 253 protonated ($-\text{NH}_3^+/\text{SH}$) (W.Y., A.B., G.L.V. and M.K., manuscript in preparation). Comparison of a quantum mechanical calculation for oxoG and G shows that the primary difference in the electrostatic potentials of the two bases is a charge inversion at positions 7 and 8 (see Fig. 4b). This creates local dipoles with opposite directions in oxoG and G. Notably, the dipole moment of the Lys 249 (NH_3^+)/Cys 253 (S^-) pair would be oriented so as to be approximately antiparallel to oxoG, but parallel to G, if G were to have a similar position as oxoG in the active site. Consequently, the Lys 249 (NH_3^+)/Cys 253 (S^-) dipole is expected to stabilize the interaction of oxoG with the active site but destabilize the interaction with G. To determine the magnitude of this effect, an alchemical simulation transforming oxoG to G was performed for the oxoG structure with the neutralized pair Lys 249 (NH_2)/Cys 253 (SH). This calculation yielded a stabilization of oxoG versus G in the oxoG active site of $5.3 \text{ kcal mol}^{-1}$, indicating that a major contribution ($3.3 \text{ kcal mol}^{-1}$) to the residue fidelity arises from the dipole–dipole interaction.

As noted above (Fig. 1a), a hydrogen bond is apparent between the carbonyl oxygen of Gly 42 and N7 H of oxoG. This attractive interaction would be replaced in the case of G by a strongly repulsive interaction with the N7 lone pair, if G and Gly 42 assumed the same positions as in the oxoG complex (Supplementary Fig. S3). A free energy perturbation analysis (see Methods) shows that the Gly 42 interaction is indeed stabilizing for oxoG versus G in the active site by about $5.3 \text{ kcal mol}^{-1}$, but this is offset in part by a destabilizing contribution of about $1.8 \text{ kcal mol}^{-1}$, due to interactions with water molecules in the active site. This result suggests that the hydrogen bond to Gly 42 makes an important contribution to stabilization of

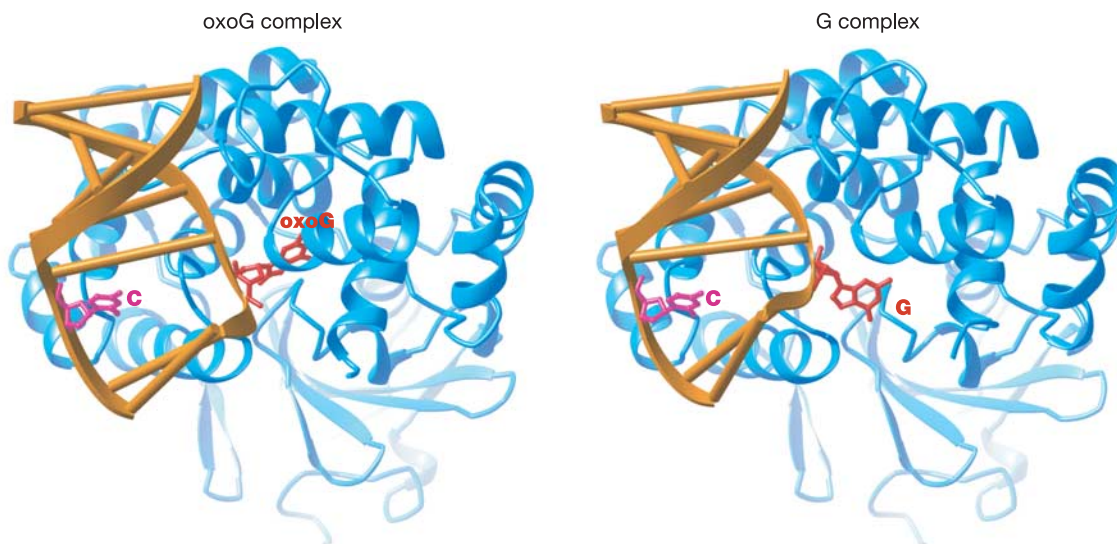


Figure 2 Comparison of the overall structures of trapped complexes obtained with oxoG-containing (left) or G-containing (right) DNA. Both protein and DNA are represented as backbone ribbon traces, with the protein in cyan and the DNA in gold. The estranged C

(magenta) and oxoG or G (red) are rendered in ball-and-stick representations. Note that oxoG is bound in the lesion recognition pocket, whereas the G is bound at the alternative extra-helical site (exo-site).

oxoG in the active site, on the order of $3.5 \text{ kcal mol}^{-1}$ (see Methods). As to the destabilizing repulsive interaction between the N7 of G and the C = O of Gly 42 (see Supplementary Fig. S3), its net contribution is less important (less than 2 kcal mol^{-1}). The reason for the small magnitude of this effect is that the hOGG1 structure is sufficiently flexible, in accord with results found more generally in simulations of proteins¹⁸, that repulsion can be relieved by a local re-orientation of Gly 42, with only a small energetic penalty (see Fig. 4c).

To explore the effect of further changes in nucleobase structure on the nature of the interactions with hOGG1, we performed alchemical simulations for two modified G analogues, 7-deazaG and 7-deaza-8-azaG (Supplementary Fig. S3), with each analogue being transmuted to G in the active and alternative sites. The results show that both 7-deazaG and 7-deaza-8-azaG, unlike G, are more stable in the active site than in the exo-site; this is in accord with the experimental structures described below. The contribution of the Lys 249 (NH_3^+)/Cys 253 (S^-) dipole is much smaller than that for oxoG, as expected; the difference in interaction energy with the ion pair and the neutral pair is less than 2 kcal. Interestingly, the major differences in $\Delta\Delta A$ for the analogues replacing oxoG arise from the poorer solvation of both analogues, relative to G, in the alternative site, where the nucleobases have significant solvent exposure.

Recognition of G analogues

To test experimentally the results of the computational studies with nucleobase analogues, we incorporated 7-deazaG into the duplex oligonucleotide, crosslinked it to hOGG1-QC, and determined the structure. As shown in Fig. 3c (see also Supplementary Fig. S4), the 7-deazaG in this structure is, like oxoG, fully inserted into the lesion recognition pocket on the enzyme. The C7 H of 7-deazaG points towards the Gly 42 carbonyl of hOGG1, and the C7 to O distance of $3.2 \text{ \AA}$ is consistent with a weak hydrogen-bonding interaction. Similar structural findings were obtained with a second analogue, 7-deaza-8-azaG, which bears a C7 H and C8 N (see Supplementary Fig. S4). These results are consistent with the computational studies described above.

Mechanism of base extrusion

A central mystery of base-excision DNA repair concerns the mechanism by which DNA glycosylases cause the extrusion of damaged bases from the genome¹⁹. The present structure sheds light on this process. The extensive structural adjustments that take place during base extrusion cannot occur in a concerted (that is, single step) process, but rather must proceed as a pathway through a series of discrete intermediates. This being the case, then the binding mode observed in the G complex could reasonably mimic a late inter-

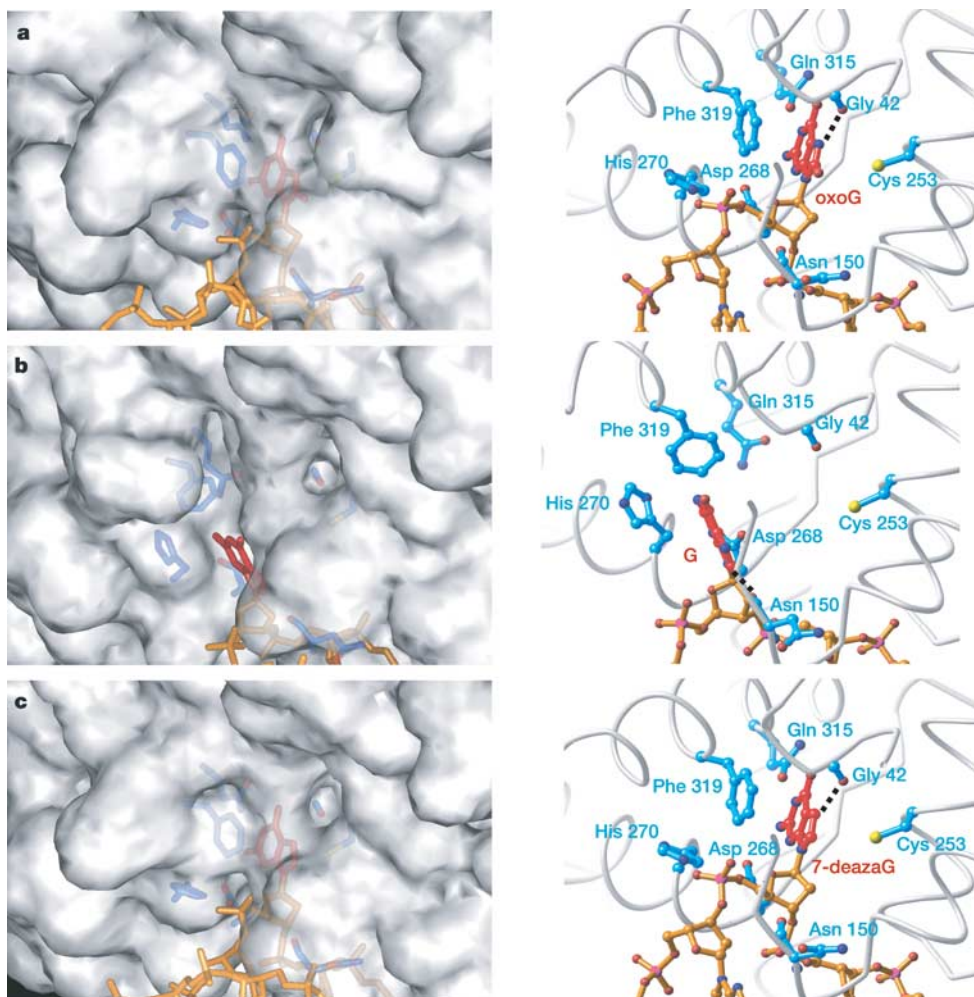


Figure 3 View of the active site region of hOGG1-DNA complexes, showing extra-helical nucleobases bound to either the lesion recognition pocket or the alternative site. **a–c**, OxoG⁴ (**a**), G (**b**) and 7-deazaG (**c**) complexes. On the left of each panel is a rendering of the solvent-accessible surface of the protein, with the DNA in framework model (DNA,

gold; oxoG/G/7-deazaG, red; key amino acid residues, blue). On the right is a detailed view of the enzyme active site, lesion recognition pocket and nearby portions of the structures (same colouring as the left panel, but with α -carbon trace of the protein backbone in grey).

mediate in the base-extrusion pathway with oxoG, just before insertion of the nucleobase into the lesion recognition pocket. Comparison of the DNA conformation in the G complex with that in the oxoG complex would thus reveal the detailed conformational changes that take place in the final stage of base extrusion, while also suggesting interactions that are established earlier in the overall extrusion pathway. Shown in Fig. 5a is a least-squares superposition of the two crosslinked complexes, using the protein component only to determine the superposition. The two cross-linked duplexes show high concordance on the 3' side of the extra-helical nucleoside (left flank), with the backbone of that strand being held in place through hydrogen-bonding contacts to the signature helix-hairpin-helix motif (Gly245, Gln249 and Val250), plus a conserved electrostatic interaction with a divalent metal ion. One exception to this concordance is the 3' phosphate of the extra-helical nucleoside, which is hydrogen-bonded to Lys249 in the G complex (predicted independently in the calculated structures); conversely, in a catalytic complex containing oxoG, Lys249 would have to be disengaged from the 3' phosphate and swung into the active site in order to fulfil its essential role in catalysis⁶. We propose that the contacts to the left-flank—perhaps even the 3' phosphate–Lys249 contact—are established early in the base-extrusion pathway, probably before the target base is actually extruded. On the other hand, the helix conformation at the position of the extra-helical nucleoside and on its right flank is markedly different in the two structures. Consequently, in the G complex the DNA has a more pronounced bend ($\sim 80^\circ$ compared with $\sim 70^\circ$), and the duplex on right flank is also over-rotated by $\sim 20^\circ$ about the helix axis (see Fig. 5c). Close inspection of the backbone conformation reveals that the difference results almost entirely from torsional adjustments in the extra-helical nucleoside and its 3' and 5' phosphates, and is associated with the loss of hydrogen-

bonding contacts to these phosphates by the Asn150 backbone NH and the side-chain imidazolium NH of His270. A divalent Ca^{2+} ion that coordinates the 3' phosphate and stabilizes the bend by inner- and outer-sphere contacts to the bases flanking the extra-helical nucleoside is also absent in the G complex (Fig. 5a). We propose that contacts to the right flank are established late in the base-extrusion pathway, only after the oxoG has achieved proper insertion into the lesion recognition pocket.

Previous comparisons of the structures of the hOGG1 lesion recognition complex⁴ and free hOGG1 (unliganded) have revealed distinct conformational states for each²⁰. Interestingly, in the G complex, the portions of hOGG1 that interact with the left flank are in the same conformation as that found in the lesion recognition complex⁴ and the oxoG complex (this work), whereas residues on the right flank and in the lesion recognition pocket, specifically His270, Phe319 and Gln315, are in the same distinctive conformation as in the free enzyme. Again, this is consistent with the notion that establishment of the left-flank contacts precedes establishment of the right-flank contacts.

On the basis of the structural comparison shown in Fig. 5 and the foregoing discussion, the final stage of the base-extrusion pathway can readily be envisioned. With the left-flank interactions firmly engaged and the nucleobase positioned on the periphery of the active site, simple bond rotations about the DNA backbone at the site of the extra-helical nucleotide slot the nucleobase into the lesion recognition pocket, while relieving over-bending and over-rotation, and enabling the formation of stabilizing backbone interactions. Large rotations ($>110^\circ$) about only three bonds drive the insertion process (Fig. 5c), facilitated by smaller adjustments of flanking bonds. Only after these events have taken place, and oxoG has entered the active site, does catalysis of base excision take place. G fails to attain this final state, primarily because of the destabilizing

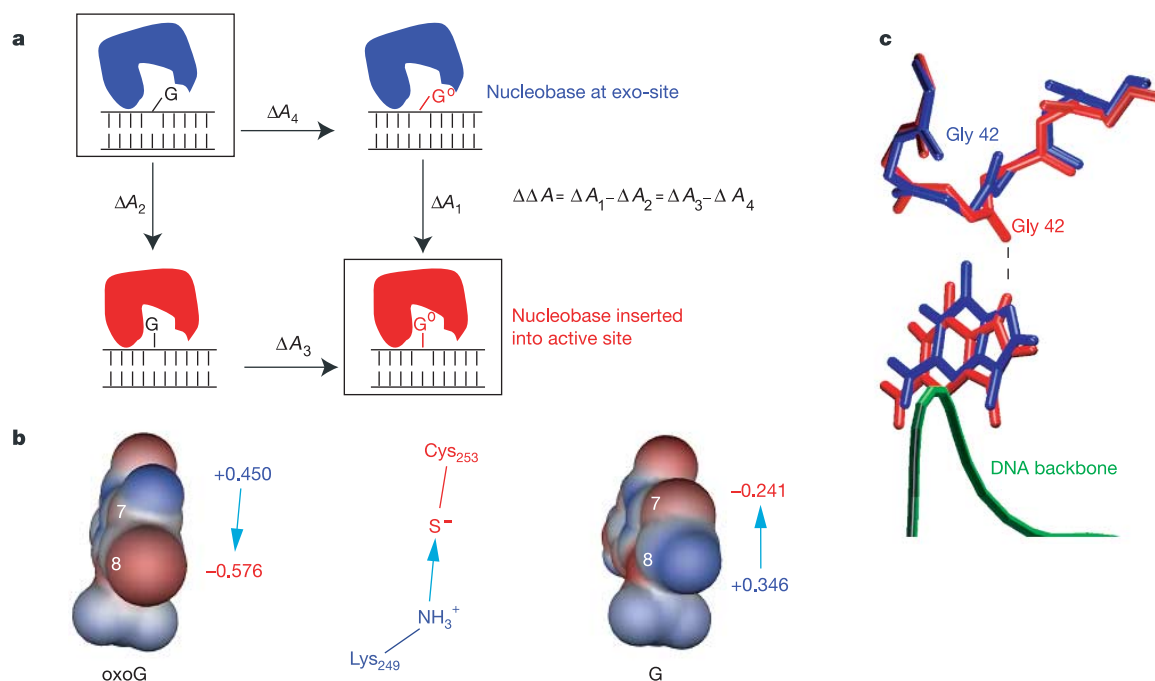


Figure 4 Computational analysis of the binding free energy difference between oxoG and G. **a**, Thermodynamic cycle for the alchemical free energy simulation of the relative binding of oxoG versus G at two binding sites. The structure containing a nucleobase at the exo-site (G versus oxoG) is schematically represented in blue, whereas that containing a nucleobase (G versus oxoG) in the active site pocket is in red. The two crystal structures used for the simulations are enclosed by a box. **b**, Electrostatic potential

difference between oxoG and G. Blue, regions of positive charge; red, regions of negative charge. Dipoles are in cyan, with Mülliken charges indicated. **c**, Orientations of the peptide group of Gly 42 in the presence of oxoG (red) versus G (blue) in the active site. The local rotations avoid a repulsive interaction between N7 of G and the carbonyl oxygen of Gly 42.

interactions it would suffer if inserted into the lesion recognition pocket; the other three bases are presumably not presented to the active site, because of the strong preference for a C at the estranged position.

The present results demonstrate that hOGG1 is capable of

discriminating G from oxoG once the nucleobase is extra-helical (extra-helical discrimination). It remains to be determined whether hOGG1 also possesses some mechanism of distinguishing oxoG from G before extrusion from the helix (intra-helical discrimination). From a biological standpoint, the complete rejection of G from the hOGG1 active site represents a gate-keeping strategy to prevent spurious cleavage of normal bases from the genome. In this regard, hOGG1 is apparently more fastidious than its paralogue 3-methyladenine DNA glycosylase (AlkA), which does occasionally excise adenine residues from undamaged DNA^{21,22}. □

Methods

hOGG1 preparation

A polymerase chain reaction fragment of hOGG1 containing amino acids 12–327 of the human *OGG1* gene was cloned into pET30a. Megaprimer mutagenesis was performed and all new constructs were fully sequenced. Expression and purification of the wild type and N149C and N149C/K249Q hOGG1 were as described⁶.

DNA preparation, disulphide crosslinking and crystallization

Phosphoramidite derivatives of oxoG, 7-deazaG, 7-deaza-8-azaG (PPG) and O⁴-triazolyl-U were purchased from Glen Research. Oligomers 5'-AGCGTCCAXGTCTACC-3', where X denotes 8-oxoG, 7-deaza-8-azaG, 7-deazaG, or G were synthesized using standard methods. Oligomers 5'-TGCTAGACCTGGACGC-3' (where the underlined C is the thiol-tethered base; Fig. 1c) were synthesized and functionalized as described²³. The protein-DNA complex was formed by mixing the duplex DNA with hOGG1 in a twofold molar excess of the protein. The crosslinked complex was purified by MonoQ. Crystallization of the crosslinked complexes was performed using the hanging-droplet vapour diffusion method at 4 °C. The oxoG complex and the complexes containing 7-deazaG and 7-deaza-8-azaG were crystallized as described⁴. The G complex was crystallized by the hanging-droplet vapour diffusion method using a well solution containing 100 mM sodium cacodylate (pH 6.0), 150 mM CaCl₂ and 15–17% PEG 8000. Crystals were transferred to a cryoprotectant containing 100 mM sodium cacodylate (pH 6.0), 150 mM CaCl₂, 17% PEG 8000 and 25% glycerol and then frozen in liquid nitrogen for data collection. Full details on the data collection and structure solution are contained in the Supplementary Information. Models for the oxoG complex contain 28 nucleotides and the complexes containing 7-deazaG and 7-deaza-8-azaG contain 25 nucleotides of DNA. The model for the G complex contains 20 nucleotides of DNA (density was weak for the outermost five base pairs and A overhang). The differences in electron density on the right flank for the oxoG complex versus the G complex presumably result from their differences in the helical conformation of that region, which is packed end-to-end with a neighbouring DNA duplex in the crystal. Renderings of the X-ray structures were prepared using Ribbons²⁴ and InsightII (Accelrys).

System studied by simulation

The crystal structures of hOGG1 in complex with oxoG-containing double-stranded DNA (Protein Data Bank code 1EBM)⁴ and the structure with G bound in the exo-site were used for the simulations; Lys 249 was initially built in based on the position of Gln 249 in the oxoG structure, and it moved during the simulations to interact strongly with Cys 253. The protonation states for all the residues were determined by the continuum electrostatics method²⁵; the results indicate that His 270 is doubly protonated in both structures. Lys 249 and Cys 253 were treated as Lys 249(+) / Cys 253(–) and as Lys 249/Cys 253 to study the contribution of these two residues to the active-site binding. For the alternative site binding, Lys 249(+) / Cys 253 was used.

For the simulation of the solvation free energy difference for the single bases in aqueous solution, a methyl group was added at the N9 position of oxoG, G, 7-deaza-8-azaG and 7-deazaG.

QM/MM free energy simulations

The chaperoned quantum mechanics/molecular mechanics (QM/MM) free energy simulation technique¹⁷ was used in the present study for the consistent description of oxoG, G and G analogues. The base of interest and part of the corresponding sugar ring (C1', C2', C4', O4' and hydrogen atoms attached to them) was treated quantum mechanically with the tight-binding DFT method, SCC-DFTB²⁶. The QM/MM linker scheme used in the present study is the same as that used in the calculation on UDG²⁷. The remaining system was treated classically with CHARMM27 parameters²⁸; CHARMM27 was also used for the chaperone potential (without van der Waals and Coulomb terms for the QM atoms), which serves to obtain correct geometries at the end points of the free energy simulations. Thermodynamic integration with the BLOCK module of the CHARMM program²⁹ was used to perturb only a portion of a base by sharing the common QM region in the sugar ring (C1', H1', C2', H2', O4', C4', H4' and link atoms). Water was treated with the modified TIP3P water model²⁹. More details concerning the methodology, including the error estimates, are given in Supplementary Information.

Free energy perturbation analysis

To compute the contribution of each residue to the free energy difference, a free energy perturbation analysis was used. It is based on the same concept as the perturbation analysis for the energy in QM/MM studies³⁰; the contribution of each residue *j* to the overall free energy is determined by computing the change in the free energy derivative resulting from turning off all the interaction between residue *j* and other atoms.

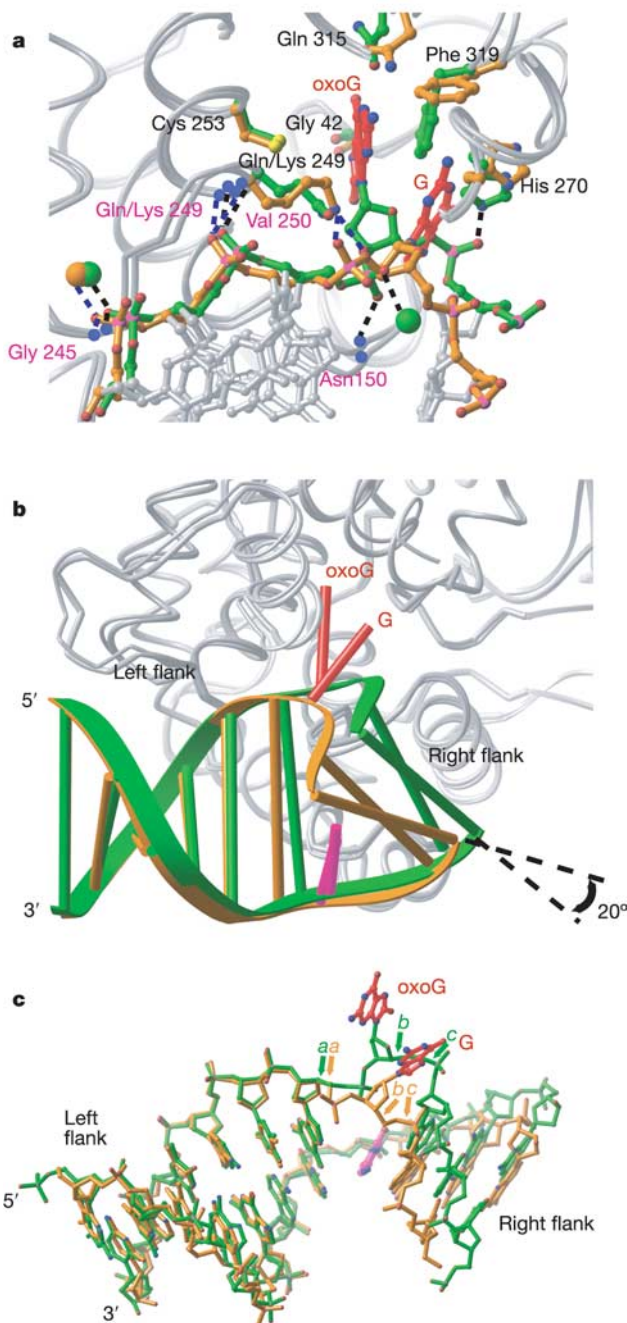


Figure 5 Superposition of the oxoG complex with the G complex in the region around the protein-DNA interface. **a**, Overlay using the protein backbone only (grey) for superposition, with the DNA backbone of the oxoG complex in green and G complex in gold. Spheres indicate Ca²⁺ ions. Residues that interact with DNA through backbone amide nitrogen atoms are denoted in magenta, whereas those that interact through side chains are in black. Dotted lines indicate hydrogen bonds. **b**, Ribbon diagram in the same orientation as **a**, but showing the whole DNA duplex. **c**, Comparison of the DNA in the two complexes, using the left flank for superposition. Arrows labelled *a*, *b* and *c* indicate bonds that have undergone significant rotations: +110° for *a* (C4'–C5' bond of the residue 3' to oxoG/G), +119° for *b* (C4'–C5' bond of oxoG/G) and –151° for *c* (P–O5' bond of oxoG/G).

Received 15 July 2004; accepted 14 February 2005; doi:10.1038/nature03458.

1. Michaels, M. L. & Miller, J. H. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J. Bacteriol.* **174**, 6321–6325 (1992).
2. Grollman, A. P. & Moriya, M. Mutagenesis by 8-oxoguanine: an enemy within. *Trends Genet.* **9**, 246–249 (1993).
3. Cappelli, E. *et al.* Rates of base excision repair are not solely dependent on levels of initiating enzymes. *Carcinogenesis* **22**, 387–393 (2001).
4. Bruner, S. D., Norman, D. P. & Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**, 859–866 (2000).
5. Fromme, J. C., Bruner, S. D., Yang, W., Karplus, M. & Verdine, G. L. Product-assisted catalysis in base-excision DNA repair. *Nature Struct. Biol.* **10**, 204–211 (2003).
6. Norman, D. P., Chung, S. J. & Verdine, G. L. Structural and biochemical exploration of a critical amino acid in human 8-oxoguanine glycosylase. *Biochemistry* **42**, 1564–1572 (2003).
7. Norman, D. P. G., Bruner, S. D. & Verdine, G. L. Coupling of substrate recognition and catalysis by a human base-excision DNA repair protein. *J. Am. Chem. Soc.* **123**, 359–360 (2001).
8. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
9. Viadiu, H. & Aggarwal, A. K. Structure of BamHI bound to nonspecific DNA: a model for DNA sliding. *Mol. Cell* **5**, 889–895 (2000).
10. Robinson, H. *et al.* The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature* **392**, 202–205 (1998).
11. Gao, Y. G. *et al.* The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. *Nature Struct. Biol.* **5**, 782–786 (1998).
12. Huang, H., Harrison, S. C. & Verdine, G. L. Trapping of a catalytic HIV reverse transcriptase•template:primer complex through a disulfide bond. *Chem. Biol.* **7**, 355–364 (2000).
13. Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669–1675 (1998).
14. Verdine, G. L. & Norman, D. P. Covalent trapping of protein-DNA complexes. *Annu. Rev. Biochem.* **72**, 337–366 (2003).
15. Fromme, J. C., Banerjee, A., Huang, S. J. & Verdine, G. L. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **427**, 652–656 (2004).
16. Simonson, T., Archontis, G. & Karplus, M. Free energy simulations come of age: protein-ligand recognition. *Acc. Chem. Res.* **35**, 430–437 (2002).
17. Yang, W., Bitetti-Putzer, R. & Karplus, M. Chaperoned alchemical free energy simulations: A general method for QM, MM, and QM/MM potentials. *J. Chem. Phys.* **120**, 9450–9453 (2004).
18. Brooks, C. L. III, Karplus, M. & Pettitt, B. M. *Advances in Chemical Physics* Vol. 71 (Wiley & Sons, New York, 1988).
19. Verdine, G. L. & Bruner, S. D. How do DNA repair proteins locate damaged bases in the genome? *Chem. Biol.* **4**, 329–334 (1997).
20. Bjoras, M., Seeberg, E., Luna, L., Pearl, L. H. & Barrett, T. E. Reciprocal “flipping” underlies substrate recognition and catalytic activation by the human 8-oxo-guanine DNA glycosylase. *J. Mol. Biol.* **317**, 171–177 (2002).

21. Berdal, K. G., Johansen, R. F. & Seeberg, E. Release of normal bases from intact DNA by a native DNA repair enzyme. *EMBO J.* **17**, 363–367 (1998).
22. O'Brien, P. J. & Ellenberger, T. The *Escherichia coli* 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site. *J. Biol. Chem.* **279**, 26876–26884 (2004).
23. MacMillan, A. M. & Verdine, G. L. Engineering tethered DNA molecules by the convertible nucleoside approach. *Tetrahedron* **47**, 2603–2616 (1991).
24. Carson, M. J. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 379–384 (1991).
25. Schaefer, M., Sommer, M. & Karplus, M. pH-dependence of protein stability: absolute electrostatic free energy differences between conformations. *J. Phys. Chem. B* **101**, 1663–1683 (1997).
26. Cui, Q., Elstner, M., Kaxiras, E., Frauenheim, T. & Karplus, M. A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B* **105**, 569–585 (2001).
27. Dinner, A. R., Blackburn, G. M. & Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* **413**, 752–755 (2001).
28. MacKerell, A. D. Jr *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
29. Simonson, T., Archontis, G. & Karplus, M. Continuum treatment of long-range interactions in free energy calculations. Application to protein-ligand binding. *J. Phys. Chem. B* **101**, 8347–8360 (1997).
30. Bash, P. A. *et al.* Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry* **30**, 5826–5832 (1991).
31. Cervi, A. R., Guy, A., Leonard, G. A., Teoule, R. & Hunter, W. N. The crystal structure of N4-methylcytosine.guanosine base-pairs in the synthetic hexanucleotide d(CGCGm4CG). *Nucleic Acids Res.* **21**, 5623–5629 (1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to Y. Korkhin for help in data collection and processing. We thank Enanta Pharmaceuticals for use of their X-ray instrumentation. We acknowledge the entire staff at MACCHESS, especially C. Heaton and B. Miller, and NSLS X4A for assistance in data collection and processing. We thank D. Jeruzalmi and C. Fromme for valuable discussions. This work was supported by grants from the NIH to G.L.V. and M.K.

Author contributions A.B. was responsible for performing the structural and biochemical experiments described herein, whereas W.Y. performed the computational simulations.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.L.V. (verdine@chemistry.harvard.edu) or M.K. (marci@tammy.harvard.edu). Atomic coordinates have been deposited in the Protein Data Bank under accession numbers 1YQK, 1YQR, 1YQL and 1YQM.

A non-terrestrial ^{16}O -rich isotopic composition for the protosolar nebula

Ko Hashizume¹ & Marc Chaussidon²

¹Department of Earth and Space Sciences, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan

²CRPG-CNRS, BP 20, 54501 Vandoeuvre-lès-Nancy Cedex, France

The discovery in primitive components of meteorites^{1,2} of large oxygen isotopic variations that could not be attributed to mass-dependent fractionation effects has raised a fundamental question: what is the composition of the protosolar gas from which the host grains formed? This composition is probably preserved in the outer layers of the Sun, but the resolution of astronomical spectroscopic measurements is still too poor to be useful for comparison with planetary material^{3,4}. Here we report a precise determination of the oxygen isotopic composition of the solar wind from particles implanted in the outer hundreds of nanometres of metallic grains in the lunar regolith. These layers of the grains are enriched in ^{16}O by $>20 \pm 4\%$ relative to the Earth, Mars and bulk meteorites, which implies the existence in the solar accretion disk of reactions—as yet unknown—that were able to change the $^{17}\text{O}/^{16}\text{O}$ and $^{18}\text{O}/^{16}\text{O}$ ratios in a way that was not dependent strictly on the mass of the isotope. Photochemical self-shielding of the CO gas irradiated by ultraviolet light^{5–7} may be one of these key processes, because it depends on the abundance of the isotopes, rather than their masses.

The isotopic composition of solar particles implanted in the first tens to hundreds of nanometres of lunar soil grains can be successfully measured by ion microprobe shallow depth profiling^{8–10}. Comparisons⁴ of such lunar-sample-based estimates^{9,10} of solar N and C isotopic compositions with recent spacecraft-based equivalents demonstrate that lunar soil samples are appropriate proxies for monitors of the solar isotopic ratios. The implanted solar O signal is anticipated to be most easily detectable in the metallic grains of the lunar regolith as they are free of structural O. We extracted 199 metallic grains and 36 sulphides by hand magnet from the Apollo 17 lunar soil sample no. 79035. Their average size is $34 \mu\text{m}$, the maximum being $113 \mu\text{m}$. Their composition was found to be either Ni-free Fe (58% of them) or Ni-bearing alloy (42%). These two types of metallic grains correspond respectively to indigenous lunar magmatic grains and to meteoritic grains delivered to the Moon by impacts¹¹. Sample 79035 is thought to have been exposed to solar radiation 1 to 2 Gyr ago⁹, and both types of grains are *a priori* suitable for the detection of implanted solar particles. We performed multi-collection depth profiling of ^{16}O , ^{17}O and ^{18}O isotopes in a subset of 38 metallic grains and 4 sulphides selected for their size and flatness (see Methods for details). An isotopic signal that can be ascribed to implanted solar O was detected in only eight grains and was most prominent in five grains (seven different profiles), which are described in the following. A typical depth profile of the O isotopic compositions and O concentrations (called here [O]) is shown for one metallic grain in Fig. 1. All the oxygen isotopic compositions and concentrations at various depths for the metallic grains are shown in Fig. 2a, b.

A surface O-rich layer (up to $\sim 24 \text{ wt}\%$, that is, close to the stoichiometry of iron oxides) of variable thickness (from 40 to 600 nm) is systematically present in all metallic grains. These oxide layers have $\Delta^{17}\text{O}$ values ($= \delta^{17}\text{O} - 0.52\delta^{18}\text{O}$), representing deviation from the terrestrial mass fractionation line (TFL), of $0.9 \pm 2.5\%$, so that they plot in Fig. 2a on the TFL. The TFL

corresponds to the terrestrial range of isotopic variations: all solar system samples should plot on the TFL if the protosolar nebula was homogenized in O isotopic composition and if no other processes than 'classical chemical and physical reactions', for which the rate constants and equilibrium constants are mass-dependent, were operating in the early Solar System. Beneath this O-rich layer, that is, for depths in excess of 500 nm, very negative $\Delta^{17}\text{O}$ values, reaching -20% at 1,200 nm depth (Fig. 1), are present. This negative $\Delta^{17}\text{O}$ signature was observed at three different points on the same grain, as well as in four other metallic grains, at depths ranging from 100 nm to $>1 \mu\text{m}$.

The $\delta^{17}\text{O}$ – $\delta^{18}\text{O}$ spread in Fig. 2a indicates the presence in the metallic grains of at least four different O-bearing components. Three of these components, which have a constant $\Delta^{17}\text{O}$ value of $\sim 0\%$ and $\delta^{18}\text{O}$ values of about -25% , $+5\%$ and $+45\%$, have been previously identified in lunar soil. The first one ($\delta^{18}\text{O} \approx -25\%$) corresponds to the surface oxide layer. Its low $\delta^{18}\text{O}$ value is consistent with an oxidation at low temperature of the metallic grains by terrestrial water¹², as suggested from its δD and $\delta^{18}\text{O}$ values plotting on the terrestrial meteoric water line. The second one ($\delta^{18}\text{O} \approx +5\%$), which is also present at the surface of some grains and which is also O-rich, corresponds clearly to lunar silicates which have $\delta^{18}\text{O}$ values^{13,14} of $+5\%$ to $+7\%$ and $\Delta^{17}\text{O}$ values^{14,15} of $0.001 \pm 0.016\%$. Scanning electron microscopy (SEM) observations of the metallic grains showed the ubiquitous presence of small silicate fragments at their surface. The third one ($\delta^{18}\text{O} \approx +45\%$) is consistent with the strongly fractionated component evidenced by partial fluorination analyses of lunar soil samples¹³. These analyses revealed that the first few per cent of O and Si released by fluorination of minerals of the soil samples were extremely mass fractionated ($\delta^{18}\text{O} \approx +50\%$ and $\delta^{30}\text{Si} \approx +25\%$), which was ascribed to some fractional vaporization/condensation of materials at the grain surfaces^{13,16}. Finally, and most importantly, the large deviation of some data in Fig. 2a from the TFL demonstrates the presence in all the grains of a fourth component with a strongly

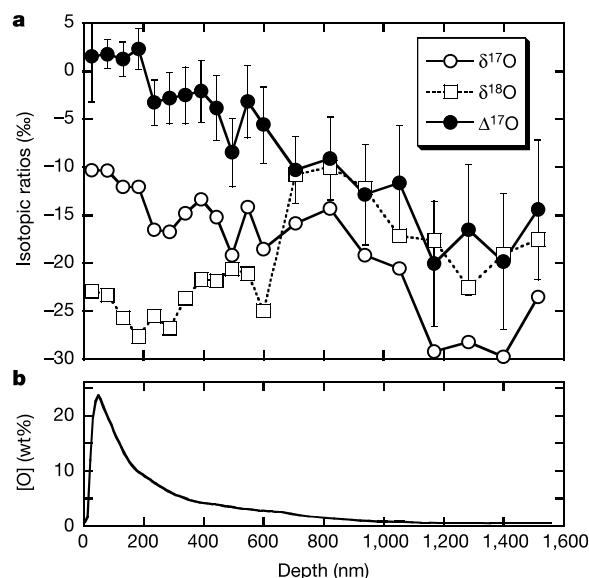


Figure 1 Typical oxygen isotope depth profiles for a lunar metallic grain showing ^{16}O enrichment. This is for spot 1 in grain 04A from soil 79035. **a**, Isotopic composition expressed in delta notation ($\delta^{17}\text{O}$ and $\delta^{18}\text{O}$) relative to the Standard Mean Ocean Water (SMOW) international standard ($\delta^{18}\text{O} = [(^{18}\text{O}/^{16}\text{O})/(^{18}\text{O}/^{16}\text{O})_{\text{SMOW}} - 1] \times 1,000$) and $\Delta^{17}\text{O}$ notation (see text) with 1σ error bars. **b**, Concentration expressed in wt%. All data for the five grains examined in this study, and typical examples for other grains, in which ^{16}O enrichments (that is, negative $\Delta^{17}\text{O}$) were not observed, are available as Supplementary Information.

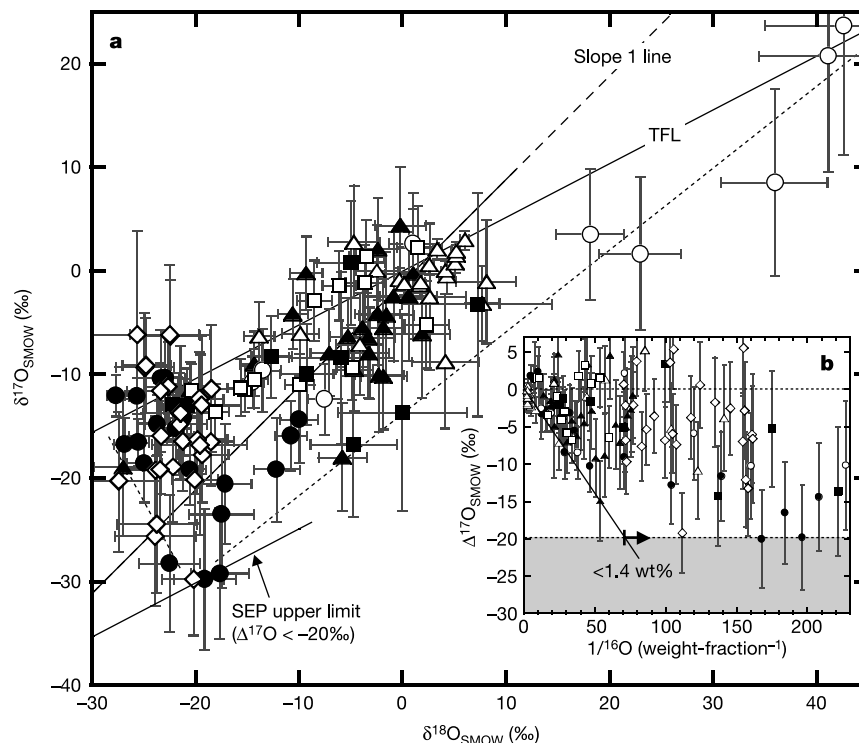


Figure 2 Oxygen isotopic composition in the surface layers of five lunar metallic grains enriched in the ^{16}O -rich SEP component. Grains are from soil 79035. Different depth profiles in grain 04A are shown as filled circles (spot 1), triangles (2) and squares (3), and profiles from grains 04B, 04C, 04D and 04E are shown as open circles, triangles, squares

and diamonds, respectively. **a**, $\delta^{17}\text{O}$ versus $\delta^{18}\text{O}$ values. **b**, Variations of $\Delta^{17}\text{O}$ values versus the inverse of the ^{16}O weight fraction. The SEP component with $\Delta^{17}\text{O} < -20\text{‰}$ has [O] of $<1.4\text{ wt\%}$ from the mixing relationships with surface components having $\Delta^{17}\text{O} \approx 0\text{‰}$. Error bars are 1σ .

negative $\Delta^{17}\text{O}$ value. It is clear from Fig. 2b that its $\Delta^{17}\text{O}$ value must be lower than or equal to $-20 \pm 4\text{‰}$ (1σ), considering the average of the three lowest $\Delta^{17}\text{O}$ values from grains 04A and 04E. We note that, from the theoretical linear mixing relationships drawn in Fig. 2a (dotted lines) between the most extreme compositions, values of $\delta^{17}\text{O} \approx -30\text{‰}$ and $\delta^{18}\text{O} \approx -20\text{‰}$ are likely for this fourth component, though these values cannot be determined precisely from the present set of data and could be even lower.

The component with a $\Delta^{17}\text{O}$ value of less than $-20 \pm 4\text{‰}$ (1σ) can be identified as implanted solar energetic particles (SEP), for the following three reasons. First, it cannot be derived by mass-dependent fractionation processes from any of the three terrestrial and lunar components mentioned above. In addition, it has no equivalent in commonly observed meteoritic components delivered to the surface of the Moon. Second, the depth at which it is observed (100 nm to $>1\text{ }\mu\text{m}$) corresponds for an implanted component to a kinetic energy of 6 to $>100\text{ keV}$ per nucleon for the incident ions¹⁷. This is higher than the typical energy of 1 keV per nucleon for the normal (or slow) solar wind (SW), and suggests that it corresponds to suprathermal solar ions such as those ubiquitously observed in the high-energy tail of the normal SW flux. Third, the mixing systematics in a $\Delta^{17}\text{O}$ versus $1/[\text{O}]$ diagram (Fig. 2b) predict that the [O] of the fourth component is lower than 1.4 wt%. This is consistent with the expected [O] of 1.0 wt% for the SEP component, calculated from the typical [C] in SEP of 5,000 p.p.m. in this sample¹⁰ and a O/C ratio in SEP¹⁸ of 2.0 g g^{-1} .

Recently, a component with a very high $\delta^{18}\text{O}$ value of approximately $+100\text{‰}$ and a high $\Delta^{17}\text{O}$ value of about $+20\text{‰}$ was found¹⁹ by ion microprobe analysis of metal particles from lunar soil sample no. 10084. A direct comparison with our results is hampered by the lack of data concerning the variation of the concentration with

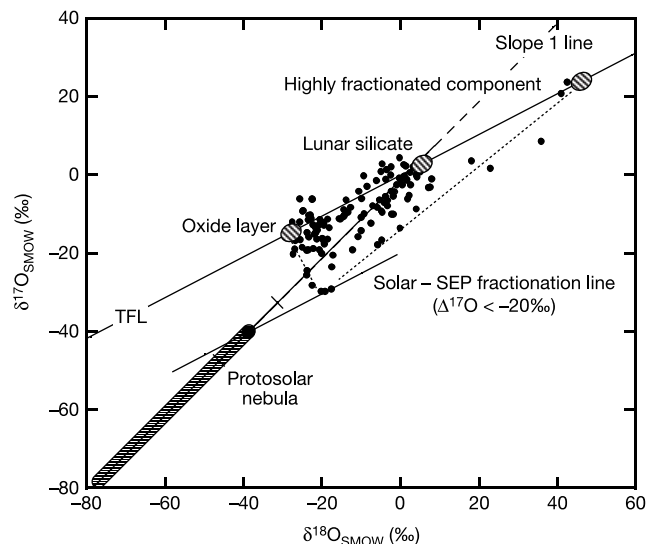


Figure 3 Prediction of the composition of the protosolar nebular gas. Theoretical mixing trajectories are shown between the SEP component and the terrestrial/lunar components identified at surfaces of the lunar metallic grains. The dots show the data points plotted in Fig. 2a. The upper limit for the predicted composition of the protosolar nebular gas, $\delta^{17}\text{O} (\approx \delta^{18}\text{O}) = -40 \pm 8\text{‰}$ (1σ), is shown at the intersection between the mass fractionation line of slope 0.52 passing through the SEP composition and the slope-1 line²² deduced from analyses of CAIs. Note that considering the carbonaceous chondrite anhydrous mineral (CCAM) line² instead has no effect on the calculated nebular gas composition.

depth of this component, which seemed however rather abundant at the surface (~ 4 wt%) of the metal. The reason for the absence of this component in the metal grains of lunar soil 79035 is not clear. We note however that sample 10084 might be a relatively young soil compared to 79035, and that several non-solar components (for example, planetary volatiles) can be expected to be more abundant in such samples²⁰.

Because the isotopic fractionations taking place during the acceleration of the SW and SEP are mass dependent²¹, the present data imply that the $\Delta^{17}\text{O}$ value of the solar atmosphere is that of SEP, that is, less than $-20 \pm 4\%$. This value is probably that of the protosolar gas to within a few per mil (ref. 4). A prediction for the $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$ values of the protosolar gas is more model-dependent, as the isotopic fractionations associated with the acceleration of the SW and SEP are not known precisely. Theoretical predictions indicate that for $\delta^{18}\text{O}$ they should be less than -50% between the solar atmosphere and the SW⁴, and less than $+260\%$ between SEP and the SW. Thus it can be predicted that in a $\delta^{17}\text{O}$ versus $\delta^{18}\text{O}$ diagram (Fig. 3) the solar atmospheric composition should lie along the left side of the mass fractionation line of slope ~ 0.52 passing through the SEP composition. Independently, the oxygen isotopic composition of the protosolar nebula plots most probably along a line with a slope of exactly one (ref. 22) defined by primary phases in Ca-Al-rich inclusions (CAIs), which are considered to be within chondrites the earliest phases condensed from the nebular gas. Thus it can be predicted from the present $\Delta^{17}\text{O}$ (less than $-20 \pm 4\%$) for the protosolar nebula that its $\delta^{17}\text{O}$ value (approximately its $\delta^{18}\text{O}$ value) is $-40 \pm 8\%$ or lower, along the slope-1 line.

A ^{16}O -rich composition for the primordial solar nebula is in agreement with observations that most refractory condensates (CAI, hibonites) in primitive meteorites^{1,22–25} are ^{16}O -rich ($\delta^{17}\text{O} \approx \delta^{18}\text{O} \approx -50\%$ to -40%), the most ^{16}O -rich composition yet found being of $\delta^{17}\text{O} \approx \delta^{18}\text{O} \approx -75\%$ in a chondrule²⁶. The fact that planetary materials have systematically lower ^{16}O abundances than refractory condensates might be explained by either (1) interactions between the ^{16}O -rich nebular gas and ^{16}O -poor presolar refractory minerals, or (2) the existence in the gas phase of non-mass-dependent reactions^{5,27,28}. The first hypothesis seems unlikely because presolar O-bearing minerals have variable $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$ values^{29,30} which do not plot on a line of slope ~ 1 passing through the protosolar nebula values. The possibility of non-mass-dependent reactions (such as those involved in the synthesis of ozone²⁷ or occurring during self-shielding of ultraviolet light by the presolar⁷ or protosolar gas^{5,6}) has been demonstrated by laboratory experiments and/or astrophysical observations, respectively. We note here that self-shielding operating in a solar or presolar CO and N₂ gas could explain, at least qualitatively, the systematic enrichments in ^{16}O , ^{12}C and ^{14}N we demonstrated for the solar gas relative to planetary materials^{9,10}. If so, most inner Solar System solids (terrestrial planets and asteroids) must have acquired their 'non-solar' oxygen isotopic compositions during condensation and isotopic exchange in a gas irradiated by ultraviolet light, which places strong constraints on the few regions of the solar accretion disk where this is possible^{5–7}. □

Methods

Ion probe analyses were performed at CRPG-CNRS (Nancy, France) using the Cameca ims-1270. Most of the technical details are similar to those previously described^{9,10}, and therefore only the outlines and the major differences compared to the previous approach are emphasized here. Metallic grains were collected from a bulk lunar soil sample (79035) by a hand magnet. No polishing or cleaning of any sort was applied to the grains. They were embedded in an indium foil and coated with gold. The grains were sputtered with a 0.5–1.5 nA primary Cs⁺ beam using electron flooding. This primary beam of 25 μm diameter was rastered over 10 $\mu\text{m} \times 10 \mu\text{m}$ to prevent contaminating oxygen residing at the surrounding surface from being transmitted to the central part (that is, the analysed part) of the rastered area. A sputtering rate of 0.7 nm s⁻¹ nA⁻¹ was determined in the metallic standards. The O isotopes were measured in multi-collection, on one Faraday cup (^{16}O), and two electron multipliers (^{17}O and ^{18}O). Because of the use of the cold trap in the

sample chamber (vacuum of $\sim 2 \times 10^{-9}$ torr) and of the generally low H content of metallic grains (implanted solar H was largely lost by volume diffusion in the present grains, which were exposed more than 1 Gyr ago and were often mildly heated after then), the intensity of $^{16}\text{OH}^-$, potentially interfering on $^{17}\text{O}^-$, was always lower or subequal to that of $^{17}\text{O}^-$. Under these conditions and with a mass resolution ($M/\Delta M$) of $\sim 6,000$, the tail of the $^{16}\text{OH}^-$ peak interfering on $^{17}\text{O}^-$ was always less than 1‰ of $^{17}\text{O}^-$. We used polished terrestrial magnetites and olivine as primary O standards, and unpolished lunar silicates as running standards. These silicate grains, which accompanied the metallic grains on hand-magnet separation, were embedded in the same indium foil, and were frequently measured between metallic grain measurements. The instrumental mass discrimination for $^{18}\text{O}/^{16}\text{O}$ ratio was about -5% . The amounts of O blank and O in the indium foil (potentially included in the sputtered area when small grains were measured) were low enough to neglect their contributions in our sample data. The two major limiting factors on the precision of the O isotopic analyses were (1) the low counting rates at depth ($^{16}\text{O}^-$ down to $\sim 5 \times 10^5$ counts s⁻¹ at the very end of a few depth profiles) in the samples, and (2) the variation of instrumental mass fractionation with orientation or flatness of the unpolished lunar grains (inducing a slight shift of the secondary ion beam pathways). The linearity and offset of the Faraday cup used to count $^{16}\text{O}^-$ was carefully monitored by analyses of silicate standards at low count rates. The uncertainty on the correction for instrumental mass fractionation can be seen in the 4‰ range measured for $\delta^{18}\text{O}$ among unpolished lunar silicate grains, which is higher than the 2‰ range reported among different lunar silicates^{13,14}. However, this uncertainty does not impinge on the determination of the $\Delta^{17}\text{O}$ values: measurements in unpolished lunar silicate grains yielded $\Delta^{17}\text{O} = +0.2 \pm 1.1\%$. Errors shown throughout the text and figures are 1 σ .

Received 24 November 2004; accepted 31 January 2005; doi:10.1038/nature03432.

1. Clayton, R. N., Grossman, L. & Mayeda, T. A component of primitive nuclear composition in carbonaceous meteorites. *Science* **182**, 485–488 (1973).
2. Clayton, R. N. Oxygen isotopes in meteorites. *Annu. Rev. Earth Planet. Sci.* **21**, 115–149 (1993).
3. Wiens, R. C., Huss, G. R. & Burnett, D. S. The solar oxygen-isotopic composition: predictions and implications for solar nebula processes. *Meteorit. Planet. Sci.* **34**, 99–107 (1999).
4. Wiens, R. C., Bochsler, P., Burnett, D. S. & Wimmer-Schweingruber, R. F. Solar and solar-wind isotopic compositions. *Earth Planet. Sci. Lett.* **222**, 697–712 (2004).
5. Clayton, R. N. Self-shielding in the solar nebula. *Nature* **415**, 860–861 (2002).
6. Lyons, J. R. & Young, E. D. Evolution of oxygen isotopes in the solar nebula. *Lunar Planet. Sci. [CD-ROM]* XXXV, 1970 (2004).
7. Yurimoto, H. & Kuramoto, K. Molecular cloud origin for the oxygen isotope heterogeneity in the solar system. *Science* **305**, 1763–1766 (2004).
8. Chaussidon, M. & Robert, F. Lithium nucleosynthesis in the Sun inferred from the solar wind $^7\text{Li}/^6\text{Li}$ ratio. *Nature* **402**, 270–274 (1999).
9. Hashizume, K., Chaussidon, M., Marty, B. & Robert, F. Solar wind record on the Moon: Deciphering presolar from planetary nitrogen. *Science* **290**, 1142–1145 (2000).
10. Hashizume, K., Chaussidon, M., Marty, B. & Terada, K. Protosolar carbon isotopic composition: implications for the origin of meteoritic organics. *Astrophys. J.* **600**, 480–484 (2004).
11. Morris, R. V. Origins and size distribution of metallic iron particles in the lunar regolith. *Proc. 11th Lunar Planet. Sci. Conf.* 1697–1712 (1980).
12. Epstein, S. & Taylor, H. P. Jr The isotopic composition and concentration of water, hydrogen and carbon in some Apollo 15 and 16 soils and in the Apollo 17 orange soil. *Proc. 4th Lunar Sci. Conf.* 1559–1575 (1973).
13. Taylor, H. P. Jr & Epstein, S. O¹⁸/O¹⁶ and Si³⁰/Si²⁸ studies of some Apollo 15, 16, and 17 samples. *Proc. 4th Lunar Sci. Conf.* 1657–1679 (1973).
14. Wiechert, U. et al. Oxygen isotopes and the Moon forming giant impact. *Science* **294**, 345–348 (2001).
15. Robert, F., Rejou-Michel, A. & Javoy, M. Oxygen isotope homogeneity of the Earth: new evidence. *Earth Planet. Sci. Lett.* **108**, 1–9 (1992).
16. Clayton, R. N., Mayeda, T. K. & Hurd, J. M. Loss of oxygen, silicon, sulfur and potassium from the lunar regolith. *Proc. 5th Lunar Sci. Conf.* 1801–1809 (1974).
17. Ziegler, J. F., Biersack, J. P. & Littmark, U. *The Stopping and Ranges of Ions in Matter* Vol. 1 (Pergamon, New York, 1985).
18. Wimmer-Schweingruber, R. F. in *Solar Wind Ten* (eds Velli, M., Bruno, R. & Malara, F.) 577–582 (AIP Press, New York, 2003).
19. Ireland, T. R., Holden, P., Norman, M. & Clarke, J. Oxygen isotopes in lunar metal grains — A natural GENESIS experiment. *Lunar Planet. Sci. [CD-ROM]* XXXV, 1448 (2004).
20. Hashizume, K., Marty, B. & Wieler, R. Analyses of nitrogen and argon in single lunar grains: towards a quantification of the asteroidal contribution to planetary surfaces. *Earth Planet. Sci. Lett.* **202**, 201–216 (2002).
21. Wieler, R. The solar noble gas record in lunar samples and meteorites. *Space Sci. Rev.* **85**, 303–314 (1998).
22. Young, E. D. & Russell, S. S. Oxygen reservoirs in the early solar nebula inferred from an Allende CAI. *Science* **282**, 452–455 (1998).
23. Krot, A. N., McKeegan, K. D., Leshin, L. A., MacPherson, G. J. & Scott, E. R. D. Existence of an ^{16}O rich gaseous reservoir in the solar nebula. *Science* **295**, 1051–1054 (2002).
24. McKeegan, K. D., Leshin, L. A., Russell, S. S. & MacPherson, G. J. Oxygen isotopic abundances in calcium-aluminum-rich inclusions from ordinary chondrites: implications for nebular heterogeneity. *Science* **280**, 414–418 (1998).
25. Ireland, T. R., Zinner, E., Fahey, A. J. & Esat, T. M. Evidence for distillation in the formation of HAL and related hibonite inclusions. *Geochim. Cosmochim. Acta* **56**, 2503–2520 (1992).
26. Kobayashi, S., Imai, H. & Yurimoto, H. New extreme ^{16}O -rich reservoir in the early solar system. *Geochim. J.* **37**, 663–669 (2003).
27. Thieme, M. H. & Heidenreich, J. E. The mass independent fractionation of oxygen: a novel isotope effect and its possible cosmochemical implications. *Science* **219**, 1073–1075 (1983).
28. Thieme, M. H. Mass-independent isotope effects in planetary atmospheres and the early solar system. *Science* **283**, 341–345 (1999).

29. Messenger, S., Keller, L. P., Stadermann, F. J., Walker, R. M. & Zinner, E. Samples of stars beyond the solar system: Silicate grains in interplanetary dust. *Science* **300**, 105–108 (2003).
30. Nagashima, K., Krot, A. N. & Yurimoto, H. Stardust silicates from primitive meteorites. *Nature* **428**, 921–924 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements Samples were provided by NASA. We thank J. Aléon, G. Libourel, B. Marty, A. Pack, F. Robert and Z. Sharp for discussions, B. Marty for encouragement, and K. Terada for help in sample preparation. This study was supported by the Mitsubishi Foundation, JSPS, MEXT, Région Lorraine, CNES-CSEEM and by INSU-PNP through a 'Poste Rouge' fellowship (K.H.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.H. (kohash@ess.sci.osaka-u.ac.jp).

Unconventional superconductivity in PuCoGa₅

N. J. Curro¹, T. Caldwell¹, E. D. Bauer¹, L. A. Morales¹, M. J. Graf¹, Y. Bang², A. V. Balatsky¹, J. D. Thompson¹ & J. L. Sarrao¹

¹Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

²Department of Physics, Chonnam National University, Kwangju 500-757, Korea

In the Bardeen–Cooper–Schrieffer theory of superconductivity, electrons form (Cooper) pairs through an interaction mediated by vibrations in the underlying crystal structure. Like lattice vibrations, antiferromagnetic fluctuations can also produce an attractive interaction creating Cooper pairs, though with spin and angular momentum properties different from those of conventional superconductors. Such interactions have been implicated for two disparate classes of materials—the copper oxides^{1,2} and a set of Ce- and U-based compounds³. But because their transition temperatures differ by nearly two orders of magnitude, this raises the question of whether a common pairing mechanism applies. PuCoGa₅ has a transition temperature intermediate between those classes and therefore may bridge these extremes⁴. Here we report measurements of the nuclear spin-lattice relaxation rate and Knight shift in PuCoGa₅, which demonstrate that it is an unconventional superconductor with properties as expected for antiferromagnetically mediated superconductivity. Scaling of the relaxation rates among all of these materials (a feature not exhibited by their Knight shifts) establishes antiferromagnetic fluctuations as a likely mechanism for their unconventional superconductivity and suggests that related classes of exotic superconductors may yet be discovered.

Cooper pairs, which have zero net spin and angular momenta in conventional superconductors, condense into a macroscopic quantum state that is separated energetically from all unpaired electrons by a finite gap Δ . Because of this gap, measurements that probe the electronic density of states near the Fermi energy exhibit a thermally activated temperature dependence below the superconducting transition temperature T_c . On the other hand, Cooper pairs formed by the exchange of antiferromagnetic spin fluctuations possess even parity and non-zero angular momentum⁵; consequently, the superconducting energy gap is not finite everywhere but vanishes at points or along lines in momentum space. These gap nodes have a profound influence on the properties observed at low temperature. For any finite temperature, well-defined electronic excitations or quasiparticles reside at these nodes, and measurements sensitive to the electronic density of states exhibit power-law variations with

temperature that depend solely on the topology of the gap zeros.

The high- T_c copper oxides and certain Ce- and U-based compounds, called heavy-fermion materials, have unconventional, nodal superconducting gaps and support antiferromagnetic fluctuations that lead naturally to these gap structures. These fluctuations common to both classes of materials are a consequence of strong electron–electron correlations: both the d -electrons in the copper oxides and the f -electrons in the heavy fermion materials experience strong on-site Coulomb repulsion that introduces elements of both localized and itinerant behaviour. In the d -electron materials these correlations create a Mott insulator in the undoped case and in the f -electron systems lead to enhanced effective electron mass. PuCoGa₅ crystallizes in exactly the same structure as does one of these unconventional, heavy-fermion superconductors: CeCoIn₅. Their common crystal structure and similarities of magnetic properties derived from nearly localized f -electrons suggest that PuCoGa₅ may also be an unconventional superconductor. However, the T_c of PuCoGa₅ is nearly an order of magnitude higher than that of CeCoIn₅ ($T_c = 2.3$ K, the highest T_c among heavy-fermion systems)⁶, but comparable to that of many conventional superconductors, such as Nb₃Sn ($T_c \approx 18$ K). Without a direct probe of the gap symmetry of PuCoGa₅ it has been impossible to differentiate conclusively between these two scenarios.

Nuclear magnetic resonance (NMR) and nuclear quadrupolar resonance (NQR) are powerful techniques used to make this distinction^{7–12}. Both techniques probe the density of quasiparticle excitations, $N(E)$, with excitation energy E above the Fermi energy, and reveal information about the spin state of the Cooper pairs and the pairing symmetry of the gap function in momentum space, $\Delta(\mathbf{k})$. We have used NMR and NQR to measure the Knight shift, K_s , and the nuclear spin-lattice relaxation rate, T_1^{-1} , of the ⁵⁹Co and ^{69,71}Ga nuclei in the normal and superconducting states of two samples of PuCoGa₅. Figure 1a shows a series of ⁷¹Ga NMR spectra obtained from Sample A (see Methods) from which the total Knight shift K_{tot} is determined straightforwardly. Similar data were collected from ⁵⁹Co NMR (not shown). The temperature dependence of these shifts, plotted in Fig. 1c, reveals a pronounced drop in K_{tot} of both ⁷¹Ga and ⁵⁹Co nuclei at T_c . In the normal states

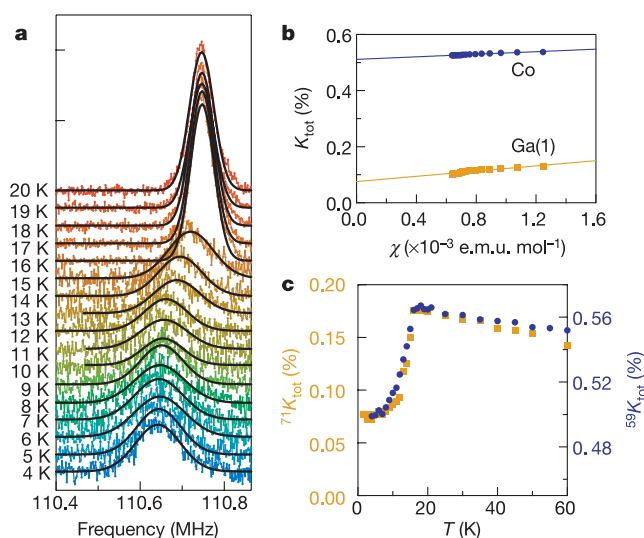


Figure 1 Knight shift measurements in PuCoGa₅. **a**, NMR spectra of ⁷¹Ga in 8 T at a series of temperatures through T_c . The spectra have been offset vertically for clarity. Solid lines are gaussian fits. **b**, The normal-state magnetic shift K_{tot} of the ⁵⁹Co and ⁷¹Ga(1) versus bulk susceptibility χ . The intercepts and hyperfine constants are given by $^{59}\text{K}_{\text{orb}} = 0.53\%$, $^{71}\text{K}_{\text{orb}} = 0.088\%$, and $^{59}\text{A} = 1.5 \text{ kOe } \mu_B^{-1}$, $^{71}\text{A} = 4.1 \text{ kOe } \mu_B^{-1}$. Solid lines are fits to the low-temperature data. **c**, The total magnetic shift K_{tot} of the ⁵⁹Co and ⁷¹Ga(1) versus temperature.

of metals like PuCoGa₅, K_{tot} is the sum of two contributions: $K_{\text{tot}} = K_{\text{orb}} + K_s$, where K_{orb} is temperature-independent, and $K_s = A\chi_s$, where A is a hyperfine constant, and χ_s is the spin susceptibility. The constants A and K_{orb} were determined independently for ⁵⁹Co and ⁷¹Ga nuclei above T_c , as shown in Fig. 1b. Quantitatively accounting for both K_{orb} and the demagnetization field in the superconducting state, we obtain the temperature dependence of χ_s shown in Fig. 2.

In the superconducting state, χ_s probes the spin symmetry of the Cooper pair wavefunction. When two quasiparticles with up- and down-spin form a Cooper pair with total spin of either $S = 0$ or $S = 1$, the wavefunction is either antisymmetric (spin-singlet or more generally even-parity) or symmetric (spin-triplet or odd-parity) under particle exchange. For singlet pairing, χ_s decreases in the superconducting state, but for triplet pairing χ_s remains constant (depending on the orientation of the applied magnetic field¹²). The decrease evident in Fig. 2 clearly establishes PuCoGa₅ as a spin-singlet superconductor. To satisfy Fermi statistics, the pair wavefunction must be antisymmetric with respect to particle exchange. Because the spin part of the wavefunction is antisymmetric (singlet), the symmetry of the orbital component of the wavefunction (given by $(-1)^L$, where the angular momentum $L = 0, 1, 2, 3, \dots$ or $s, p, d, f, \dots$) must be symmetric, so L must be even. Thus, the results of Fig. 2 leave open the possibility that PuCoGa₅ could be either a conventional s -wave or an unconventional superconductor with $L > 0$ and even.

Although χ_s drops below T_c , as expected for spin-singlet pairing, it should approach zero as $T \rightarrow 0$ (neglecting minor vortex core contributions). In an unconventional, nodal superconductor, impurity scattering creates excitations in the superconducting gap nodes, and hence a finite spin susceptibility. Impurity scattering in PuCoGa₅ is inevitable owing to lattice defects created by the recoil of uranium atoms during the radioactive decay of the plutonium, for example, $^{239}\text{Pu} \rightarrow ^{235}\text{U} + \alpha\text{-particle}$ ¹³. The solid curve in Fig. 2, discussed below, shows that the temperature dependence of χ_s is described well by calculations that account for impurity scattering in a d -wave superconductor.

Further evidence for nodal superconductivity is provided by

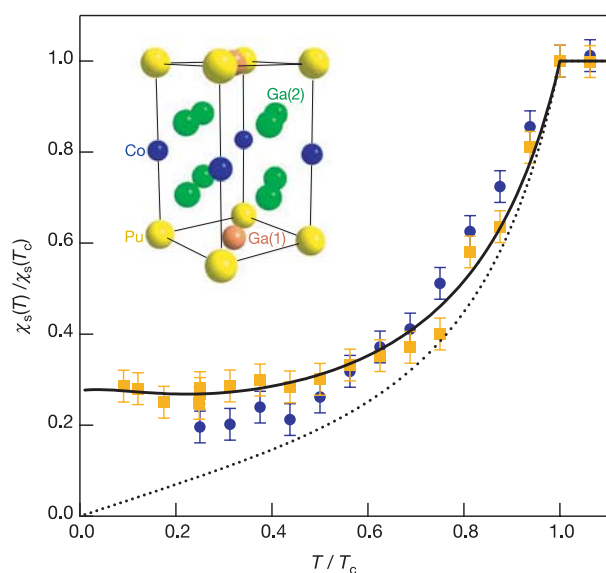


Figure 2 Normalized spin susceptibility in the superconducting state. The ⁵⁹Co (blue circles) and ⁷¹Ga(1) (orange squares) data as well as calculations for pure d -wave (dotted line) and dirty d -wave (solid line) gap functions are shown. The latter assumes strong impurity scattering in the self-consistent T-matrix approximation (SCTA) with scattering rate $\Gamma^A/\Delta = 0.03$, gap $\Delta/k_B T_c = 4$ and $b = 1.74$. Inset, the PuCoGa₅ crystal structure, with the Ga site index.

spin-lattice relaxation rate measurements on Sample B (see Methods). T_1^{-1} measures the rate at which a nucleus reaches an equilibrium spin temperature; it is dominated by scattering between the conduction electrons and the nuclear spins. In the superconducting state, the condensate cannot relax the nuclei without breaking Cooper pairs, so the nuclei are predominantly relaxed by quasiparticles. Because the symmetry of the orbital part of the Cooper pair wavefunction determines the energy dependence of $N(E)$ below the energy gap Δ , T_1^{-1} establishes the pairing symmetry. The data are shown in Fig. 3a. Below T_c , T_1^{-1} exhibits a sharp decrease, with roughly $T_1^{-1} \propto T^3$ behaviour, followed by $T_1^{-1} \propto T$ at the lowest temperatures. In contrast to expectations for an s -wave superconductor, the lack of a (Hebel–Slichter coherence) peak in T_1^{-1} just below T_c and the power-law dependence of T_1^{-1} are identical to responses found in the copper oxide superconductors and CeCoIn₅ (refs 14, 15). Furthermore, the temperature dependence of T_1^{-1} in the normal state of PuCoGa₅ is qualitatively different than that observed in conventional BCS superconductors

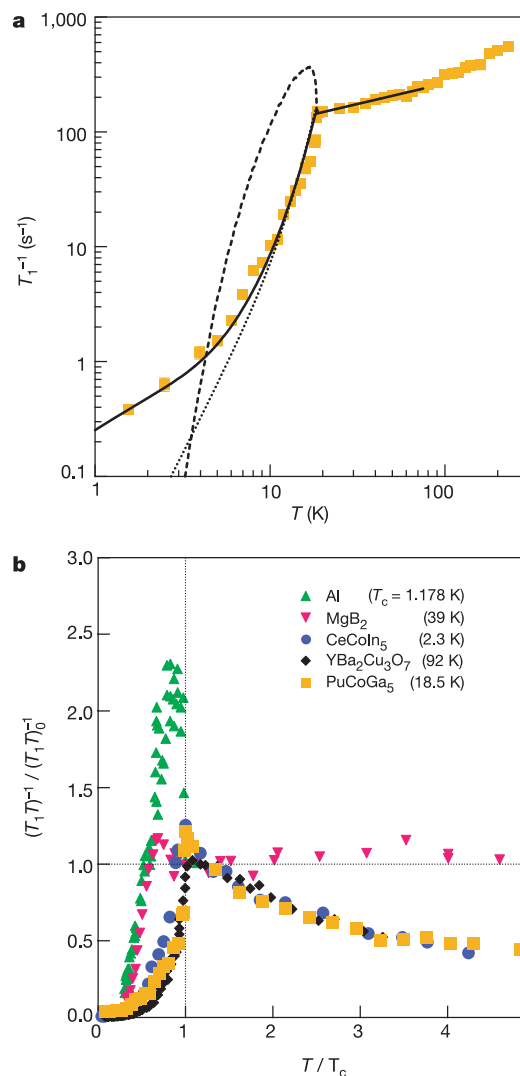


Figure 3 The spin-lattice relaxation rate in the normal and superconducting states. **a**, T_1^{-1} data for the ⁶⁹Ga(1), as well as calculations for BCS isotropic s -wave (dashed), pure (dotted) and dirty (solid) d -wave gap functions. The solid line in the normal state shows $T^{0.35}$. **b**, $(T_1 T)^{-1}/(T_1 T_0)^{-1}$ versus T/T_c for PuCoGa₅, as well as for the unconventional superconductors YBa₂Cu₃O₇ ($T_c = 92$ K; ref. 7) and CeCoIn₅ ($T_c = 2.3$ K; ref. 27), and the s -wave superconductors Al ($T_c = 1.178$ K; ref. 8) and MgB₂, ($T_c = 39.2$ K; ref. 11). The normalization constant $(T_1 T_0)^{-1}$ is given by the value of $(T_1 T)^{-1}$ at $1.25 T_c$ (see Methods).

(see Fig. 3b), but has been predicted for a relaxation rate dominated by antiferromagnetic spin fluctuations¹⁶. Further, the normalized relaxation rates of PuCoGa₅, high- T_c YBa₂Cu₃O₇ and CeCoIn₅ scale onto a common curve as a function of the dimensionless parameter T/T_c . Although we have chosen T_c as an easily defined characteristic energy, we also find that, with a spin-fluctuation energy of $T_0 \approx 270$ K for PuCoGa₅ (ref. 17), the T_c and T_0 values of PuCoGa₅ have the same proportionality observed for several other unconventional superconductors¹⁸ (see Supplementary Fig. 1). The results of Fig. 3b, consequently, argue that PuCoGa₅ is indeed a bridge between extreme cases. A priori, we might expect a similar scaling behaviour among the Knight shifts of these materials; however, the long wavelength ($q = 0$) response in the normal state of these systems is quite different from the finite q response that dominates the spin lattice relaxation² (see Supplementary Fig. 2).

Self-consistency of our results is provided by calculations of the temperature dependence of T_1^{-1} below T_c for sample B and the Knight shift for sample A within the framework of a self-consistent T-matrix approximation (SCTA)^{19,20}, considering the effects of impurity scattering on both s - and d -wave pairing scenarios. The T_1^{-1} data are best-fitted (Fig. 3a, solid curve) by a strong-coupling d -wave gap function with lines of nodes and the parameters $\Delta/k_B T_c = 4$ and $\Gamma^B/\Delta = 0.01$, where Γ^B is the impurity scattering rate for sample B. This gap value is significantly enhanced relative to the d -wave weak-coupling limit $\Delta/k_B T_c = 2.14$ and very similar to that determined for CeCoIn₅ (ref. 15). The temperature dependence of χ_s (Fig. 2 data) is fitted using the same gap magnitude, but with $\Gamma^A/\Delta = 0.03$ for sample A. The difference in the impurity lifetime broadening $\Delta\Gamma = \Gamma^A - \Gamma^B$ is due to the age difference and hence impurity scattering in samples A and B. From the Abrikosov-Gorkov relationship²¹ $\Delta T_c = (\pi/4)\Delta\Gamma$, this difference ($\Delta\Gamma/\Delta = 0.02$) implies $\Delta T_c = 1.2$ K, in agreement with both the difference in T_c values of the two samples and time-dependent studies of T_c suppression ($dT_c/dt \approx -0.24$ K per month; Jutier, F. *et al.*, unpublished work). From our fits, we estimate $d\Gamma/dt \approx 0.25$ K per month, which implies that $T_{c0} = 19.1$ K for pristine, defect-free PuCoGa₅, a value roughly half that of recent theoretical predictions²².

The total electronic energy calculations for PuCoGa₅ and its isostructural neighbour NpCoGa₅ predict that both should order antiferromagnetically²³. As predicted, NpCoGa₅ is an antiferromagnet below the Néel temperature $T_N \approx 47$ K (ref. 24), but there is no evidence for long-range magnetic order above 1 K in PuCoGa₅ (ref. 25). These calculations neglect the role of many-electron correlations that lead to a nearly magnetic state in Ce- and U-based heavy-fermion systems, which, in the absence of such correlations, would order magnetically. The existence of these correlations is predicted for Pu-based materials²⁶ and is reflected in PuCoGa₅ through its enhanced electronic specific heat coefficient, $\gamma \approx 95$ mJ mol⁻¹ K⁻² (ref. 17). As with the copper oxides and heavy-fermion systems like CeCoIn₅, the Cooper pairing in PuCoGa₅ is most probably mediated by antiferromagnetic fluctuations arising from proximity to an antiferromagnetic/paramagnetic border.

It thus appears that PuCoGa₅ establishes continuity in the spectrum of energy scales controlling unconventional superconductivity in other f -electron systems and in the copper oxides, and that a universal tunable magnetic pairing mechanism may be operative in all materials with functional elements of both localized and itinerant electrons. This leads naturally to the speculation that there may be other materials classes in which antiferromagnetic fluctuations create an exotic form of superconductivity. □

Methods

Spin lattice relaxation and Knight shift analysis

There are several nuclei in PuCoGa₅ that can be investigated by NMR: ⁶⁹Ga ($I = 3/2$), ⁷¹Ga ($I = 3/2$), and ⁵⁹Co ($I = 7/2$). The nuclear hamiltonian is given by

$\hat{H} = \gamma \hbar \mathbf{B}_0 (1 + K_{\text{tot}}) + \hbar \nu_c [3\hat{I}_z^2 - \hat{I}^2] - \eta (\hat{I}_x^2 - \hat{I}_y^2)$, where h is Planck's constant, γ is the gyromagnetic ratio, $\mathbf{B}_0$ is the external magnetic field, $\nu_c = eQV_{\text{cd}}/20$, $\eta = (V_{\text{aa}} - V_{\text{bb}})/V_{\text{cc}}$, Q is the quadrupolar moment and the V values are the components of the electric field gradient tensor. In the PuCoGa₅ structure (Fig. 2 inset) (space group $P4/mmm$), the Co (1b site) lies directly below and above each Pu, has axial symmetry, and experiences an electric field gradient given by $^{59}\nu_Q = 1.70$ MHz, $\eta = 0$. There are two different ^{69,71}Ga sites: Ga(1) (1c site) and Ga(2) (4i site). The Ga(1) site has axial symmetry, has four nearest-neighbour Pu atoms, and experiences a large electric field gradient: $^{69}\nu_Q = 28.28$ MHz, $\eta = 0$. For the Knight shift measurements, both the central line ($I_z = -1/2 \leftrightarrow +1/2$) of the Co and the upper satellite of the ⁷¹Ga ($I_z = +3/2 \leftrightarrow +1/2$) were measured. The resonance frequencies f of these transitions can be written in second-order perturbation theory (because $\gamma \mathbf{B}_0 \gg \nu_Q$) by:

$$^{59}f = ^{59}\gamma \mathbf{B}_0 (1 + K) + 15 \times ^{59}\nu_Q^2 / (16 \times ^{59}\gamma \mathbf{B}_0) [1 - 10 \cos^2(\theta) + 9 \cos^4(\theta)] \quad (1)$$

$$^{71}f = ^{71}\gamma \mathbf{B}_0 (1 + K) + ^{71}\nu_Q (1 - 3 \cos(2\theta)) / 4 + 3 \times ^{71}\nu_Q^2 / (2 \times ^{71}\gamma \mathbf{B}_0) \cos^2(\theta) \sin^2(\theta) \quad (2)$$

where θ is the angle between $\mathbf{B}_0$ and the c axis of the crystal. T_1^{-1} was measured at the zero-field quadrupolar transition ($I_z = \pm 3/2 \leftrightarrow \pm 1/2$) of the ⁶⁹Ga(1) by fitting the time dependence of the magnetization recovery after inversion of the nuclear polarization. Sample A, used to measure K_s , was a single crystal oriented with the c axis 74.3° from the applied field (8 T) and was measured six months after its growth. Because of radiation damage, the T_c of sample A was reduced from its as-grown value of 18.5 K and in the measuring field of 8 T was ~ 16 K. Sample B, used to measure T_1^{-1} , consisted of an unoriented powder that had aged two weeks, with $T_c \approx 18.5$ K. Each sample was encapsulated inside an NMR solenoid coil embedded in Stycast epoxy. Thermal contact to the sample was established via gas transfer through stainless steel frits with 2-μm pore sizes located along the axis of the coil.

In the superconducting state, there is a third contribution to the total shift: $-\Delta B/B_0$, where ΔB (the demagnetization field) measures the reduction in the internal magnetic induction due to the diamagnetic shielding currents induced by the applied field. By measuring two different sites (⁵⁹Co and ⁷¹Ga(1)) in PuCoGa₅, we determine $\Delta B \approx 40$ Oe at 4 K (ref. 9). Because of the field orientation in our experiment, we can rule out a spin-triplet pairing state with a $\mathbf{d}$ -vector pointing along the crystallographic c axis, as proposed for the spin-triplet superconductors Sr₂RuO₄ and UPt₃ (ref. 12).

For the theoretical fits we modelled the temperature dependence of the strong-coupling superconducting gap function as $\Delta T = \Delta \tanh[b(T_c/T - 1)^{0.5}]$. The ratio of the specific-heat jump to C_N (the normal-state specific heat at T_c) constrained the phenomenological parameters Δ and b as follows: $\Delta C/C_N = (b\Delta/(\pi k_B T_c))^2/(2a)$, with $a = 2/3$ for a free-electron gas. Our choice of parameters $\Delta/k_B T_c = 4$ and $b = 1.74$ lead to a moderately enhanced coefficient $a \approx 1.1$ –1.7 for $\Delta C/C_N \approx 1.43$ –2.28, consistent with reported values for the Sommerfeld coefficient $\gamma_s = C_N/T$ for $T \rightarrow 0$ (refs 4, 23).

The spin lattice relaxation and spin susceptibility data were fitted to:

$$T_1^{-1}(T)/T_1^{-1} = -2(T/T_c) \int_0^\infty dE [N^2(E) + M^2(E)]/N_F^2 df(E)/dE \quad (3)$$

$$\chi_s(T)/\chi_n = -2(T/T_c) \int_0^\infty dE N(E)/N_F df(E)/dE \quad (4)$$

where $f(E)$ is the Fermi-Dirac function, N_F is the density of states at E_F , T_n^{-1} is the normal state relaxation rate at T_c , and $M(E)$ is the 'anomalous' density of quasiparticle states, which vanishes for d -wave superconductors. The 'dirty' d -wave calculation (assuming a finite density of states from impurities) in Fig. 3a assumes strong impurity scattering in the SCTA with $\Gamma^B/\Delta = 0.01$, $\Delta/k_B T_c = 4$ and $b = 1.74$. Note that it is not possible to fit both the suppression of the Hebel-Slichter coherence peak just below T_c and the linear low-temperature behaviour with an isotropic s -wave pairing model, even when magnetic impurity scattering is included.

The normal-state relaxation rate data shown in Fig. 3b have been normalized to the value at $1.25T_c$ in order to avoid complications associated with the pseudogap effect, which suppresses T_1^{-1} in the normal state just above T_c in the copper oxides¹⁴. There is a qualitative difference between the conventional and the d -wave superconductors below and above T_c . For the d -wave superconductors, the relaxation rate is given by a single scaling function $f(T/T_c)$ up to at least $3T_c$.

Received 3 September 2004; accepted 4 February 2005; doi:10.1038/nature03428.

- Monthoux, P. *et al.* Toward a theory of high-temperature superconductivity in the antiferromagnetically correlated cuprate oxides. *Phys. Rev. Lett.* **67**, 3448–3451 (1991).
- Moriya, T. & Ueda, K. Spin fluctuation spectra and high temperature superconductivity. *J. Phys. Soc. Jpn* **63**, 1871–1880 (1994).
- Mathur, N. D. *et al.* Magnetically mediated superconductivity in heavy fermion compounds. *Nature* **394**, 39–43 (1998).
- Sarrao, J. L. *et al.* Plutonium-based superconductivity with a transition temperature above 18 K. *Nature* **420**, 297–299 (2002).
- Anderson, P. W. Further consequences of symmetry in heavy-electron superconductors. *Phys. Rev. B* **32**, 499 (1985).
- Petrovic, C. *et al.* Heavy-fermion superconductivity in CeCoIn₅ at 2.3 K. *J. Phys. Condens. Matter* **13**, L337–L342 (2001).
- Ohsugi, S. *et al.* Nuclear relaxation in strong coupling superconductors — a comparison with high- T_c superconductors. *J. Phys. Soc. Jpn* **61**, 3054–3057 (1992).
- Masuda, Y. & Redfield, A. G. Nuclear spin-lattice relaxation in superconducting aluminium. *Phys. Rev.* **125**, 159–163 (1962).
- Stenger, V. A. *et al.* Nuclear magnetic resonance of A₃C₆₀ superconductors. *Phys. Rev. Lett.* **74**, 1649–1652 (1995).
- Ueda, K. *et al.* ²⁹Si Knight shift in the heavy-fermion superconductor CeCu₂Si₂. *J. Phys. Soc. Jpn* **56**, 867–870 (1987).

11. Kotegawa, H. *et al.* Evidence for strong-coupling *s*-wave superconductivity in MgB_2 : ^{11}B NMR study. *Phys. Rev. Lett.* **87**, 127001 (2001).
12. Ishida, K. *et al.* Spin-triplet superconductivity in Sr_2RuO_4 identified by ^{17}O Knight shift. *Nature* **396**, 658–660 (1998).
13. Wolfer, W. G. Radiation effects in plutonium. *Los Alamos Sci.* **26**, 274–285 (2000).
14. Kitaoka, Y. *et al.* Nuclear relaxation and Knight shift studies of copper in $\text{YBa}_2\text{Cu}_3\text{O}_{7-y}$. *J. Phys. Soc. Jpn* **57**, 30–33 (1988).
15. Kohori, Y. *et al.* NMR and NQR studies of the heavy fermion superconductors CoTiIn_5 ($T = \text{Co}$ and Ir). *Phys. Rev. B* **64**, 134526 (2001).
16. Ishigaki, A. & Moriya, T. Nuclear magnetic relaxation around the magnetic instability in metals. *J. Phys. Soc. Jpn* **65**, 3402–3403 (1996).
17. Bauer, E. D. *et al.* Structural tuning of unconventional superconductivity in PuMGa_5 ($M = \text{Co, Rh}$). *Phys. Rev. Lett.* **93**, 147005 (2004).
18. Moriya, T. & Ueda, K. Antiferromagnetic spin fluctuations and superconductivity. *Rep. Prog. Phys.* **66**, 1299–1341 (2003).
19. Bang, Y. *et al.* T_{C}^{-1} in the *d*-wave superconducting state with coexisting antiferromagnetism. *Phys. Rev. B* **69**, 014505 (2004).
20. Hirschfeld, P. J. *et al.* Consequences of resonant impurity scattering in anisotropic superconductors: thermal and spin properties. *Phys. Rev. B* **37**, 83–97 (1988).
21. Abrikosov, A. A. *et al.* *Methods of Quantum Field Theory in Statistical Physics* Ch. 7, Sec. 39.3 (Dover, New York, 1975).
22. Bang, Y. *et al.* Possible pairing mechanisms of PuCoGa_5 superconductor. *Phys. Rev. B* **70**, 104512 (2004).
23. Opahle, I. & Oppeneer, P. M. Superconductivity caused by the pairing of plutonium 5*f* electrons in PuCoGa_5 . *Phys. Rev. Lett.* **90**, 157001 (2003).
24. Colineau, E. *et al.* Magnetic and electronic properties of the antiferromagnet NpCoGa_5 . *Phys. Rev. B* **69**, 184411 (2004).
25. Griveau, J. C. *et al.* Pressure dependence of the superconductivity in PuCoGa_5 . *J. Magn. Magn. Mater.* **272/276**, 154–155 (2004).
26. Savrasov, S. Y. & Kotliar, G. Ground state theory of δ -Pu. *Phys. Rev. Lett.* **84**, 3670–3673 (2000).
27. Kawasaki, Y. *et al.* Anisotropic spin fluctuations in heavy-fermion superconductor CeCoIn_5 : In-NQR and Co-NMR studies. *J. Phys. Soc. Jpn* **72**, 2308–2311 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank Z. Fisk, D. Pines and C. P. Slichter for discussions. This work was performed at Los Alamos National Laboratory under the auspices of the US Department of Energy Office of Science. Y.B. is supported by KOSEF through CSCMR.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.J.C. (curro@lanl.gov).

Two-dimensional spectroscopy of electronic couplings in photosynthesis

Tobias Brixner¹, Jens Stenger¹, Harsha M. Vaswani¹, Minhaeng Cho², Robert E. Blankenship³ & Graham R. Fleming¹

¹Department of Chemistry, and the Institute for Quantitative Biomedical Research (QB3), University of California, Berkeley, and Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

²Department of Chemistry and Center for Multidimensional Spectroscopy, Division of Chemistry and Molecular Engineering, Korea University, Seoul 136-701, Korea

³Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287, USA

Time-resolved optical spectroscopy is widely used to study vibrational and electronic dynamics by monitoring transient changes in excited state populations on a femtosecond timescale¹. Yet the fundamental cause of electronic and vibrational dynamics—the coupling between the different energy levels involved—is usually inferred only indirectly. Two-dimensional femtosecond infrared spectroscopy based on the heterodyne detection of three-pulse photon echoes^{2–7} has recently allowed the direct mapping of vibrational couplings, yielding transient

structural information. Here we extend the approach to the visible range^{3,8} and directly measure electronic couplings in a molecular complex, the Fenna–Matthews–Olson photosynthetic light-harvesting protein^{9,10}. As in all photosynthetic systems, the conversion of light into chemical energy is driven by electronic couplings that ensure the efficient transport of energy from light-capturing antenna pigments to the reaction centre¹¹. We monitor this process as a function of time and frequency and show that excitation energy does not simply cascade stepwise down the energy ladder. We find instead distinct energy transport pathways that depend sensitively on the detailed spatial properties of the delocalized excited-state wavefunctions of the whole pigment–protein complex.

The Fenna–Matthews–Olson (FMO) bacteriochlorophyll *a* (BChl) protein of green sulphur bacteria serves both as an antenna for collecting light energy and as a mediator for directing light excitations from the chlorosome antenna to the reaction centre^{9–12}. The FMO protein is a trimer of identical subunits, each containing seven BChl pigments. Because of its comparatively simple structure it is often employed as a model system for excitonic (delocalization) effects in photosynthesis research¹³. However, it is not clear to what extent individual spectral features, such as those in the linear FMO absorption spectrum, are caused by differences in site energies (arising from the interaction of the BChl pigments with their local protein environment) or by energy splitting from excitonic coupling (that is, pigment–pigment interaction)^{14–16}. With two-dimensional femtosecond spectroscopy, couplings can be made visible directly.

Two-dimensional (2D) optical spectroscopy^{3,8,17–21} measures the full nonlinear polarization of a quantum system in third order with respect to the field–matter interaction^{3,8,22}. A sequence of three ultrashort laser pulses excites the sample, and the emitted phase-matched signal field is detected in amplitude and phase as a function of optical frequency and the three experimentally controlled excitation–pulse time delays τ , T and t . Fourier transforms of the data^{8,21} reveal complex-valued 2D (ω_τ, ω_t) frequency maps of the system's response, where ω_τ and ω_t are the conjugate Fourier frequencies of τ and t . The real part can be interpreted as the transient field amplitude at a particular probe frequency ω_p , induced by a specific excitation frequency ω_τ after the waiting time (population time) T . The imaginary part describes transient changes in the refractive index. Whereas diagonal peaks in the 2D traces correspond to linear absorption positions, any off-diagonal contributions indicate coupling and, for $T > 0$, energy transfer. In the infrared regime, 2D maps for vibrational couplings have been determined^{2,4–7}, yielding transient structural information. Two-colour photon-echo spectroscopy has been used to study electronic coupling in a homodimer²³, but molecular cross-peak features for electronic transitions in the visible range have not yet been reported.

The experimental 2D trace of the FMO complex for $T = 0$ fs is shown in Fig. 1a. The positions of the three main diagonal peaks match the linear absorption spectrum below (Fig. 1d, solid black line). Elongation of these features along the diagonal indicates a correlation between excitation and emission frequencies within the same pigment and hence inhomogeneous spectral broadening^{2,24,25}. Analysis of the 2D contour shapes can recover the homogeneous linewidths. More importantly, however, several off-diagonal features (positive and negative) are clearly visible. Using an assignment of exciton levels (horizontal and vertical lines numbered 1–7 in Fig. 1a), one can qualitatively identify the extent of mutual correlations as the magnitude of the corresponding cross peaks. Quantitative evaluation is done by comparison with simulations as shown below. The cross-peak amplitude is determined by quantum-mechanical interference from different nonlinear optical transition pathways. Because of the electronic coupling between pigments, the excitonic transition dipoles become linear combinations of pigment

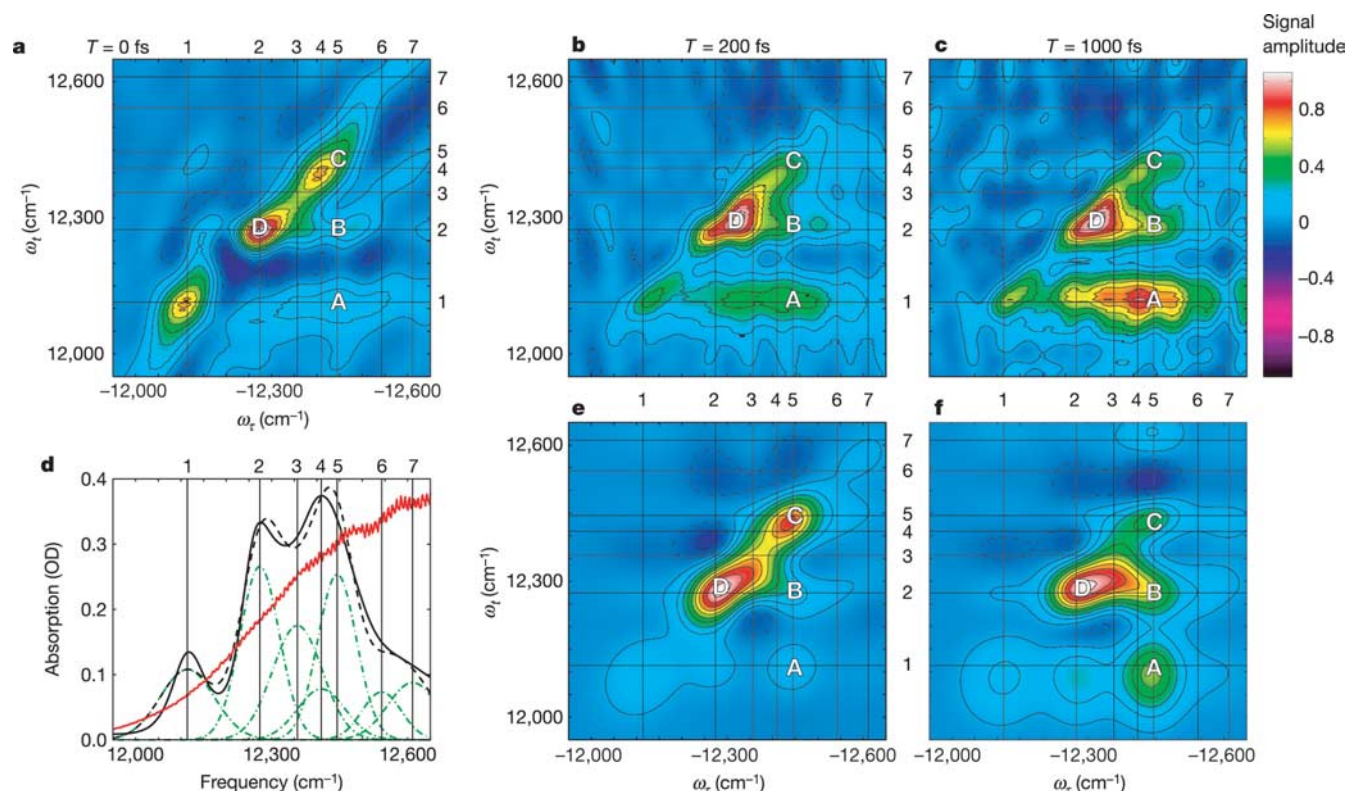


Figure 1 Experimental and theoretical spectra (real parts) of the FMO complex from *Chlorobium tepidum* at 77 K. **a–c**, The experimental 2D spectra (upper three panels) are shown for population times $T = 0$ fs (**a**), $T = 200$ fs (**b**) and $T = 1,000$ fs (**c**). Contour lines are drawn in 10% intervals at $\pm 5\%$, $\pm 15\%$, ..., $\pm 95\%$ of the peak amplitude, with solid lines representing positive contributions ('more light') and dashed lines negative features. Horizontal and vertical grid lines indicate exciton levels 1–7 as labelled. **d**, The

experimental linear absorption spectrum (solid black) is reproduced by theory (dashed black), with individual exciton contributions as shown (dashed–dotted green). The laser spectrum used in the 2D experiments (red) covers all transition frequencies. **e, f**, Simulations of 2D spectra are shown for $T = 200$ fs (**e**) and $T = 1,000$ fs (**f**). Off-diagonal features such as those labelled A and B are indicators of electronic coupling and energy transport; they are discussed in more detail in the text.

transition dipoles, and nonlinear optical transitions involving two different exciton states are allowed and produce off-diagonal peaks even at $T = 0$. The transitions above and below the diagonal involve different one-exciton and two-exciton states, with correspondingly different oscillator strengths. Accordingly, the cancellation between bleaching/stimulated emission features and excited-state absorption features is different in the two half-planes, giving rise to the stronger cross peaks below the diagonal. For cross peak A, the system is in an electronic coherence state of the ground and fifth excitonic states during τ and of the lowest one-exciton and ground states during t . This cross peak indicates that the BChl pigments that make up excitons 1 and 5 are coupled; feature B shows the same for excitons 2 and 5. The negative regions (dark blue in Fig. 1a) can be attributed to two-exciton contributions, that is, excited-state absorption.

The electronic couplings lead to energy transfer, which is observed for population times $T > 0$. As an example, we show two snapshots at $T = 200$ fs (Fig. 1b) and $T = 1,000$ fs (Fig. 1c). With these 2D spectra, not only do we measure the population evolution as in conventional pump–probe experiments, we also follow the state-to-state energy transfer pathway. Between 200 and 1,000 fs, the amplitude of the lowest-energy diagonal peak increases and the main diagonal peak (D) shifts to lower energies, indicating a sizeable downward population transfer. The concentration of features below the diagonal also indicates that 'downhill' transfer, from higher to lower energies, dominates. The cross peaks in Fig. 1b, c demonstrate the sensitivity of this method to pigment–pigment interactions. Focusing on the two dominant off-diagonal peaks near regions A and B, the 2D spectra show increasing amplitudes as

a function of waiting time T . This reveals, for example, the relaxation from exciton states 4 and 5 to exciton state 2 (B) and exciton state 1 (A) as a function of time. Other features can be discussed qualitatively in an analogous fashion.

For quantitative simulations we use a Frenkel exciton hamiltonian with electronic coupling constants and site energies obtained by fitting the resulting absorption and 2D spectra. A single ohmic spectral density (representing the chromophore–bath coupling-strength distribution) and modified Förster/Redfield theory^{26,27} are employed for fully self-consistent calculations of the exciton transfer rates, the linear absorption spectrum (Fig. 1d, dashed black line) and the time-dependent 2D spectra (Fig. 1e, f). By using modified Förster/Redfield theory, we can self-consistently describe the excited states as either localized or delocalized (excitonic) depending on the magnitude of the coupling constants²⁷. Most of the exciton states are delocalized over two or three molecules, but the lowest exciton is essentially localized on BChl 3 (refs 14, 28). The transport of excitation is treated as reversible; finite temperature and detailed balance are correctly incorporated. The comparison between experiment and theory shows that the positions of the peaks in the 2D frequency space are reproduced. In addition, the relative timescales associated with the appearance/disappearance of the specific features are also well described. For example, in both theory and experiment the amplitudes in cross peaks A and B increase with population time T as a result of downhill energy transfer. The experimental (and theoretical) ratios of the amplitude of cross peak B to diagonal peaks C and D at $T = 1$ ps are 1.49 (1.45) and 0.55 (0.5), respectively. The calculated relative amplitude of diagonal peak C to D (0.34) at $T = 1$ ps is also close to that from

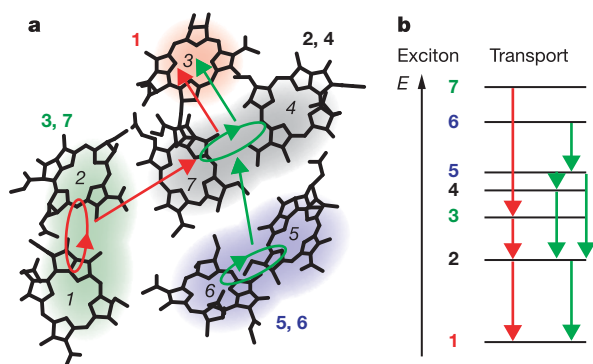


Figure 2 Exciton delocalization and energy transport. **a**, The FMO structural arrangement of the seven BChl molecules (italic numbers) is overlaid qualitatively with the delocalization patterns of the different excitons (coloured shading, bold numbers). Two main photoexcitation transfer pathways are indicated by red and green arrows. **b**, The energy transport is not just a simple process of stepwise energy decrease from one level to the next level below; rather, intermediate states are left out if they have insufficient spatial overlap with potential transfer partners.

experiment (0.37). We reproduce the presence of negative regions (dashed contour lines in Fig. 1c, f) corresponding to excited-state absorption to the two-exciton states, as well as the shape-contour change around the central diagonal peak, that is, its horizontal elongation towards region B. Some differences between experiment and simulation also occur. First, the build-up of cross peak A is slower in the simulations (Fig. 1f) than in experiment (Fig. 1c), implying that the calculated rate of populating level 1 is too slow; that is, the ratios of the amplitude of cross peak A to that of diagonal peaks C and D in the experimental (theoretical) spectra at $T = 1$ ps are 2.07 (1.41) and 0.76 (0.48), respectively. Second, the experimental cross peak near A (Fig. 1c) extends horizontally between exciton states 2 and 5, whereas in the simulation (Fig. 1f) separate peaks appear. This is at least partly the result of our limited (40 cm^{-1}) ω_T -frequency resolution.

Nevertheless, considering the complexity of the physical system, the agreement with the key 2D features is very promising. On the basis of our model, we now analyse how the excitation propagates between the different BChl molecules; that is, we follow the energy transfer in space and time. First we consider the delocalization of the exciton wavefunctions. Most excitons are delocalized mainly over only one or two neighbouring BChl molecules. This is indicated by the coloured shading in Fig. 2a. For example, excitons 3 and 7 (green, bold numbers) are both delocalized over the same BChls 1 and 2 (italic numbers). Second, from the exciton transfer-rate matrix obtained with the modified Förster/Redfield theory we can then identify the two major energy transport pathways shown by red and green arrows. In terms of energy levels, the results are displayed in Fig. 2b. An important factor for efficient energy transport is the mutual wavefunction overlap between initial and final states. Stronger overlap leads to faster transfer.

Consider first the 'green' pathway, starting with exciton 6 (on BChls 5 and 6). Because of its strong wavefunction overlap with exciton 5 (located on the same two pigments), the excitation is transferred efficiently (green ellipse). From there, exciton 4 can be reached, which in turn transfers to exciton 2 (again because of the strong spatial overlap as both excitons 2 and 4 are located on BChls 4 and 7). Alternatively, exciton 2 is reached directly from exciton 5. Note that in either case exciton 3 is not involved because it does not have strong spatial overlap with exciton 4 or 5. From exciton 2, the transport proceeds to exciton 1, and ultimately to the reaction centre where it is turned into chemical energy. The second pathway (red arrows) starts with exciton 7 and proceeds through excitons 3

and 2 to 1. Using similar arguments, it is clear that the good spatial overlap between excitons 7 and 3 and the weak coupling of corresponding BChls 1 and 2 to the remaining pigments prevents energy transfer from exciton 7 to excitons 4, 5 or 6. We conclude that the energy is not simply transferred stepwise down the energy ladder as has been conjectured previously^{15,28}. Instead, distinct pathways emerge (Fig. 2b) in which some energetically intermediate states are left out.

As illustrated by these findings, 2D femtosecond photon-echo spectroscopy remedies the deficiency of conventional types of spectroscopy in which only the evolutions of populations over time can be determined directly, but not the mechanisms that underlie these temporal changes. Thus we get detailed insight into the driving force of biological light harvesting, and applications to larger photosynthetic systems are possible. By noting that the magnitudes of electronic couplings and exciton relaxation rates are essentially dependent on the spatio-energetic distribution of chromophores, analysis of the exciton delocalization pattern leads to spatial information on the molecular scale. Hence, in general, the combination of these fully self-consistent calculations and experiments allows us to follow energy transport through space and time with nanometre spatial resolution and femtosecond temporal resolution. This methodology also opens the door to similar investigations of electronic couplings and energy transport in any photoactive assembly, macromolecule or other nanoscale system. □

Methods

Experiment and data analysis

Our implementation of inherently phase-stabilized two-dimensional Fourier-transform femtosecond spectroscopy for electronic transitions has been described in detail elsewhere^{20,21}. In brief, 50-fs, 805-nm, 3-kHz laser pulses from a home-built Tisapphire regenerative-amplifier laser system are used for three-pulse photon-echo spectroscopy in a diffractive-optic-based^{29,30} non-collinear four-wave mixing set-up with phase-matched box geometry. Time delays are introduced with better than $\lambda/100$ precision by movable glass wedges, and passive interferometric phase stability is maintained over several hours^{20,21}. The third-order signal is completely characterized by spectral interferometry with the help of a heterodyning local-oscillator pulse¹⁷. Automated subtraction of scattering terms removes experimental artefacts caused by sample imperfections. To resolve individual spectral features, we perform the experiment at low temperature (77 K) in a liquid-nitrogen cryostat (Oxford Instruments).

For any given population time (waiting time) T , the coherence time τ is scanned in 5-fs steps from -440 fs to $+440$ fs, moving excitation pulse 1 (2) in the second (first) half of the scanning period. Spectral interferograms are recorded with a 16-bit, 100-pixel $\times$ 1,340-pixel, thermoelectrically cooled charge-coupled device camera (Princeton Instruments) and a 0.3-m imaging spectrometer (Acton). These parameters lead to a spectral resolution of 40 cm^{-1} for ω_T and 2 cm^{-1} for ω_r . The spot diameter at the sample position is $84 \mu\text{m}$ ($1/e^2$ intensity level), and the excitation energy is 20 nJ per pulse for the 2D traces shown. Repetitions with 30% of the laser power led to essentially the same results within the experimental uncertainties (though at decreased signal-to-noise ratio). Data analysis by Fourier transformation yields the desired 2D traces whose absolute phase is obtained by means of the projection-slice theorem and comparison with separately recorded spectrally resolved pump-probe data²¹. Each 2D trace is the average of three separate scans.

The FMO complex of *Chlorobium tepidum* was prepared in a buffer of 50 mM Tris HCl at pH 8.0 and 10 mM sodium ascorbate¹⁰. To avoid cracking of samples at low temperature, we used a water/glycerol (35:65 v/v) mixture between plastic windows (thickness 0.3 mm) with a 0.4 mm optical path length. The optical density peak value was 0.37 at 77 K. After each series of 2D scans for different population times, we repeated the first of the scans in the same sample spot for comparison. This gave qualitatively the same 2D results within the experimental noise limits as those shown, and the absolute decrease in 2D signal due to sample degradation was below 30%.

Theory and numerical simulations

The Frenkel exciton hamiltonian matrix elements, denoted by H_{jk} , were obtained by simultaneously fitting to the absorption and time-resolved 2D photon-echo spectra. The diagonal elements, H_{jj} , for BChl molecules $j = 1, \dots, 7$ (Fig. 2a), are 280, 420, 0, 175, 320, 360 and 260 cm^{-1} . The off-diagonal matrix elements are identical to those presented previously¹⁴, except that $H_{56} = H_{65}$ was reduced to 40 cm^{-1} because the corresponding diagonal frequencies were readjusted in the present work. The general nonlinear response functions were obtained^{3,22}, but the Laplace (stationary-phase approximation) method was used to calculate the corresponding 2D Fourier spectrum approximately. The site energy fluctuation was assumed to be uncorrelated with those of other sites. The frequency-frequency correlation function determining line broadening processes was expressed in terms of an ohmic-type spectral density, that is, $\rho(\omega) = (\lambda/\hbar\omega_c)\omega^{-1}\exp(-\omega/\omega_c)$, where

the cutoff frequency ω_c and the solvent reorganization energy λ were assumed to be 40 and 35 cm^{-1} , respectively. To calculate the rate constants we used modified Förster/Redfield theory^{26,27}. The rate constants are fully determined by the spectral density defined above and the electronic coupling constants. Here, the cutoff frequency dividing Förster and Redfield regimes was assumed to be 30 cm^{-1} so that if a given pair of pigments were coupled to each other weakly (less than 30 cm^{-1}) the excitation transfer process was treated as a Förster process. Denoting the exciton transport rate constant from the j th one-exciton state to the i th one-exciton state by k_{ij} , we found that $k_{21} = 0.12$, $k_{21} = 2.97$, $k_{13} = 0.12$, $k_{23} = 0.28$, $k_{24} = 5.38$, $k_{25} = 1.6$, $k_{26} = 0.22$, $k_{35} = 0.17$, $k_{37} = 1.6$, $k_{42} = 0.48$, $k_{45} = 2.0$, $k_{46} = 2.46$, $k_{54} = 0.91$, $k_{56} = 5.73$, $k_{57} = 0.16$, $k_{64} = 0.18$, $k_{65} = 0.92$, $k_{67} = 0.68$ and $k_{76} = 0.22$ (all in ps^{-1}). All other rate constants are smaller than 0.1 ps^{-1} . The initial populations of the seven exciton states were determined by using the experimental laser spectrum (Fig. 1d, red).

Received 29 November 2004; accepted 4 February 2005; doi:10.1038/nature03429.

1. Zewail, A. H. *Femtochemistry* (World Scientific, Singapore, 1994).
2. Asplund, M. C., Zanni, M. T. & Hochstrasser, R. M. Two-dimensional infrared spectroscopy of peptides by phase-controlled femtosecond vibrational photon echoes. *Proc. Natl Acad. Sci. USA* **97**, 8219–8224 (2000).
3. Mukamel, S. Multidimensional femtosecond correlation spectroscopies of electronic and vibrational excitations. *Annu. Rev. Phys. Chem.* **51**, 691–729 (2000).
4. Wright, J. C. Coherent multidimensional vibrational spectroscopy. *Int. Rev. Phys. Chem.* **21**, 185–255 (2002).
5. Khalil, M., Demirdöven, N. & Tokmakoff, A. Coherent 2D IR spectroscopy: Molecular structure and dynamics in solution. *J. Phys. Chem. A* **107**, 5258–5279 (2003).
6. Asbury, J. B. *et al.* Hydrogen bond dynamics probed with ultrafast infrared heterodyne-detected multidimensional vibrational stimulated echoes. *Phys. Rev. Lett.* **91**, 237402 (2003).
7. Cervetto, V., Helbing, J., Bredenbeck, J. & Hamm, P. Double-resonance versus pulsed Fourier transform two-dimensional infrared spectroscopy: An experimental and theoretical comparison. *J. Chem. Phys.* **121**, 5935–5942 (2004).
8. Jonas, D. M. Two-dimensional femtosecond spectroscopy. *Annu. Rev. Phys. Chem.* **54**, 425–463 (2003).
9. Fenna, R. E. & Matthews, B. W. Chlorophyll arrangement in a bacteriochlorophyll protein from *Chlorobium limicola*. *Nature* **258**, 573–577 (1975).
10. Li, Y.-F., Zhou, W., Blankenship, R. E. & Allen, J. P. Crystal structure of the bacteriochlorophyll *a* protein from *Chlorobium tepidum*. *J. Mol. Biol.* **271**, 456–471 (1997).
11. Blankenship, R. E. *Molecular Mechanisms of Photosynthesis* (Blackwell, Oxford, 2002).
12. Blankenship, R. E. & Matsuura, K. in *Light-Harvesting Antennas in Photosynthesis* (eds Green, B. R. & Parson, W. W.) 195–217 (Kluwer Academic, Dordrecht, 2003).
13. Savikhin, S., Buck, D. R. & Struve, W. S. Toward level-to-level energy transfers in photosynthesis: The Fenna–Matthews–Olson protein. *J. Phys. Chem. B* **102**, 5556–5565 (1998).
14. Vulto, S. I. E. *et al.* Exciton simulations of optical spectra of the FMO complex from the green sulfur bacterium *Chlorobium tepidum* at 6 K. *J. Phys. Chem. B* **102**, 9577–9582 (1998).
15. Vulto, S. I. E. *et al.* Excited state dynamics in FMO antenna complexes from photosynthetic green sulfur bacteria: A kinetic model. *J. Phys. Chem. B* **103**, 8153–8161 (1999).
16. Wendling, M. *et al.* The quantitative relationship between structure and polarized spectroscopy in the FMO complex of *Prosthecochloris aestuarii*: Refining experiments and simulations. *Photosynth. Res.* **71**, 99–123 (2002).
17. Lepetit, L. & Joffe, M. Two-dimensional nonlinear optics using Fourier-transform spectral interferometry. *Opt. Lett.* **21**, 564–566 (1996).
18. Tian, P., Keusters, D., Suzuki, Y. & Warren, W. S. Femtosecond phase-coherent two-dimensional spectroscopy. *Science* **300**, 1553–1555 (2003).
19. Cowan, M. L., Ogilvie, J. P. & Miller, R. J. D. Two-dimensional spectroscopy using diffractive optics based phased-locked photon echoes. *Chem. Phys. Lett.* **386**, 184–189 (2004).
20. Brixner, T., Stopkin, I. V. & Fleming, G. R. Tunable two-dimensional femtosecond spectroscopy. *Opt. Lett.* **29**, 884–886 (2004).
21. Brixner, T., Mancal, T., Stopkin, I. V. & Fleming, G. R. Phase-stabilized two-dimensional electronic spectroscopy. *J. Chem. Phys.* **121**, 4221–4236 (2004).
22. Cho, M. Nonlinear response functions for three-dimensional spectroscopies. *J. Chem. Phys.* **115**, 4424–4437 (2001).
23. Prall, B. S., Parkinson, D. Y., Fleming, G. R., Yang, M. & Ishikawa, N. Two-dimensional optical spectroscopy: Two-color photon echoes of electronically coupled phthalocyanine dimers. *J. Chem. Phys.* **120**, 2537–2540 (2004).
24. Tokmakoff, A. Two-dimensional line shapes derived from coherent third-order nonlinear spectroscopy. *J. Phys. Chem. A* **104**, 4247–4255 (2000).
25. Kwac, K. & Cho, M. Two-color pump-probe spectroscopies of two- and three-level systems: Two-dimensional line shapes and solvation dynamics. *J. Phys. Chem. A* **107**, 5903–5912 (2003).
26. Zhang, W. M., Meier, T., Chernyak, V. & Mukamel, S. Exciton-migration and three-pulse femtosecond optical spectroscopies of photosynthetic antenna complexes. *J. Chem. Phys.* **108**, 7763–7774 (1998).
27. Yang, M., Damjanovic, A., Vaswani, H. M. & Fleming, G. R. Energy transfer in photosystem I of cyanobacteria *Synechococcus elongatus*: Model study with structure-based semi-empirical Hamiltonian and experimental spectral density. *Biophys. J.* **85**, 140–158 (2003).
28. van Amerongen, H., Valkunas, L. & van Grondelle, R. *Photosynthetic Excitons* (World Scientific, Singapore, 2000).
29. Maznev, A. A., Nelson, K. A. & Rogers, T. A. Optical heterodyne detection of laser-induced gratings. *Opt. Lett.* **23**, 1319–1321 (1998).
30. Goodno, G. D., Dadusc, G. & Miller, R. J. D. Ultrafast heterodyne-detected transient-grating spectroscopy using diffractive optics. *J. Opt. Soc. Am. B* **15**, 1791–1794 (1998).

Acknowledgements We thank Y.-Z. Ma, L. Valkunas and M. Yang for discussions, and C. Goodhope for protein purification. The apparatus for 2D spectroscopy was constructed by I. V. Stiopkin and T.B. This work was supported by the DOE (at LBNL, UC Berkeley and Arizona State University), and by a CRIP grant to M.C. by KOSEF (Korea). T.B. thanks the German Science Foundation (DFG) for an Emmy Noether fellowship, and J.S. thanks the German Academic Exchange Service (DAAD) for a postdoctoral fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.R.F. (grfleming@lbl.gov) and M.C. (mcho@korea.ac.kr).

Decline of the marine ecosystem caused by a reduction in the Atlantic overturning circulation

Andreas Schmittner

College of Oceanic and Atmospheric Sciences, Oregon State University, 104 COAS Admin. Bldg, Corvallis, Oregon 97331, USA

Reorganizations of the Atlantic meridional overturning circulation were associated with large and abrupt climatic changes in the North Atlantic region during the last glacial period^{1–4}. Projections with climate models suggest that similar reorganizations may also occur in response to anthropogenic global warming^{5–7}. Here I use ensemble simulations with a coupled climate–ecosystem model of intermediate complexity to investigate the possible consequences of such disturbances to the marine ecosystem. In the simulations, a disruption of the Atlantic meridional overturning circulation leads to a collapse of the North Atlantic plankton stocks to less than half of their initial biomass, owing to rapid shoaling of winter mixed layers and their associated separation from the deep ocean nutrient reservoir. Globally integrated export production declines by more than 20 per cent owing to reduced upwelling of nutrient-rich deep water and gradual depletion of upper ocean nutrient concentrations. These model results are consistent with the available high-resolution palaeorecord, and suggest that global ocean productivity is sensitive to changes in the Atlantic meridional overturning circulation.

The climatic consequences of Atlantic meridional overturning (AMO) reorganizations have been intensely documented in recent decades^{1–4}. Temperature oscillations in Greenland, recorded in the isotopic composition of the ice, show rapid warmings of about 10°C , known as Dansgaard–Oeschger events, coincident with abrupt increases of sea surface temperature (SST) and sea surface salinity (SSS) in the North Atlantic. During the cold (stadial) phases of the oscillations, deep water formation in the North Atlantic was partially or entirely stopped³, leading to reduced northward heat transport by the ocean. Observed spatial patterns, amplitudes and phasing of temperature changes associated with Dansgaard–Oeschger oscillations can be successfully reproduced with coupled climate models forced with disturbances of the Atlantic freshwater budget⁴. However, the consequences of AMO changes for the marine ecosystem have not yet been quantified on a global scale. This will be particularly important, as projections of future climate change indicate that the AMO could weaken or even disappear in the coming centuries owing to anthropogenic greenhouse gas emissions^{5–7}. Here I examine the impact of changes in ocean circulation on the marine ecosystem.

A climate model of intermediate complexity is used, including

four different formulations of simple upper ocean ecosystem components (see Methods section for a more detailed description and references). In order to trigger transitions between different states of the AMO, a slowly varying freshwater perturbation to the North Atlantic (Fig. 1a) is applied, following ref. 4. This leads to a gradual spindown of the circulation from about 16 Sv at year 0 to less than 3 Sv at year 500. Associated with this spindown is a reduction of convection and deep water formation and strong cooling (not shown here; see, for example, ref. 4) in the North Atlantic. This represents the transition from an interstadial (warm) to a stadial (cold) phase of the Dansgaard-Oeschger oscillations. The climate system remains in the stadial mode for more than 800 yr until around year 1,340 when a rapid return to interstadial conditions occurs.

After the AMO spindown, global export production (through the sinking of particulate organic matter), as well as primary production and biomass, slowly decrease by up to 10–20% around year 1,200 (Fig. 1b–d). The decrease in productivity is caused by a reduction of the upper ocean nutrient concentrations (Fig. 1e). Export production in the Atlantic rapidly reduces by 40% until year 500 (Fig. 1f). In the Indian and Pacific oceans, productivity decline is slower and reaches one-third around years 1,000 and 1,200, respectively. In the Southern Ocean, relative changes are smaller (10%) than in the other basins.

The response is most dramatic in the North Atlantic. Integrated plankton stocks north of 35° N plunge by $51 \pm 8\%$ from year 0 to year 500 (Fig. 1g). Annual net primary productivity in most of the

North Atlantic collapses (Fig. 2d). Nutrient concentrations in today's North Atlantic are relatively small (Fig. 3a, c). However, winter mixed layers are deep, allowing replenishment of surface nutrient levels from quite large depths. Shoaling of mixed layers (Figs 1i, 3e) as a consequence of the AMO reduction and the associated development of a strong halocline decreases nutrient supply to the photic zone, leading to rapid plankton demise.

After deep water formation has ceased, nutrient levels slowly increase in the North Atlantic (Fig. 1h) owing to horizontal advection and diffusion along isopycnals, and a shallow nutricline develops (Fig. 3e), similar to the present situation in the North Pacific (Fig. 3b, d). This is in good agreement with glacial reconstructions suggesting higher nutrient levels in the deep and intermediate North Atlantic during cold phases^{2,8}. (Note that carbon isotope ratios used as nutrient proxies in refs 2 and 8 might also be influenced by air-sea gas exchange.) In the top 1 km of the Pacific, Indian and South Atlantic oceans, nutrient levels decrease while deep sea nutrient concentrations increase owing to missing injection of low-nutrient North Atlantic Deep Water (Fig. 3). Deep upwelling in the Indo-Pacific is reduced by 50% (4 Sv), contributing directly to lesser nutrients there. Upwelling across 1,000 m into the Southern Ocean (south of 35° S) diminishes by 10 Sv, causing lower nutrient concentrations in the formation areas of Subantarctic Mode Water (SAMW). Large regions of the world's upper oceans are currently supplied with nutrients by SAMW⁹. Thus, lower nutrient levels of SAMW reduce nutrient supply to these regions, as analysis of the individual term balances of the upper ocean

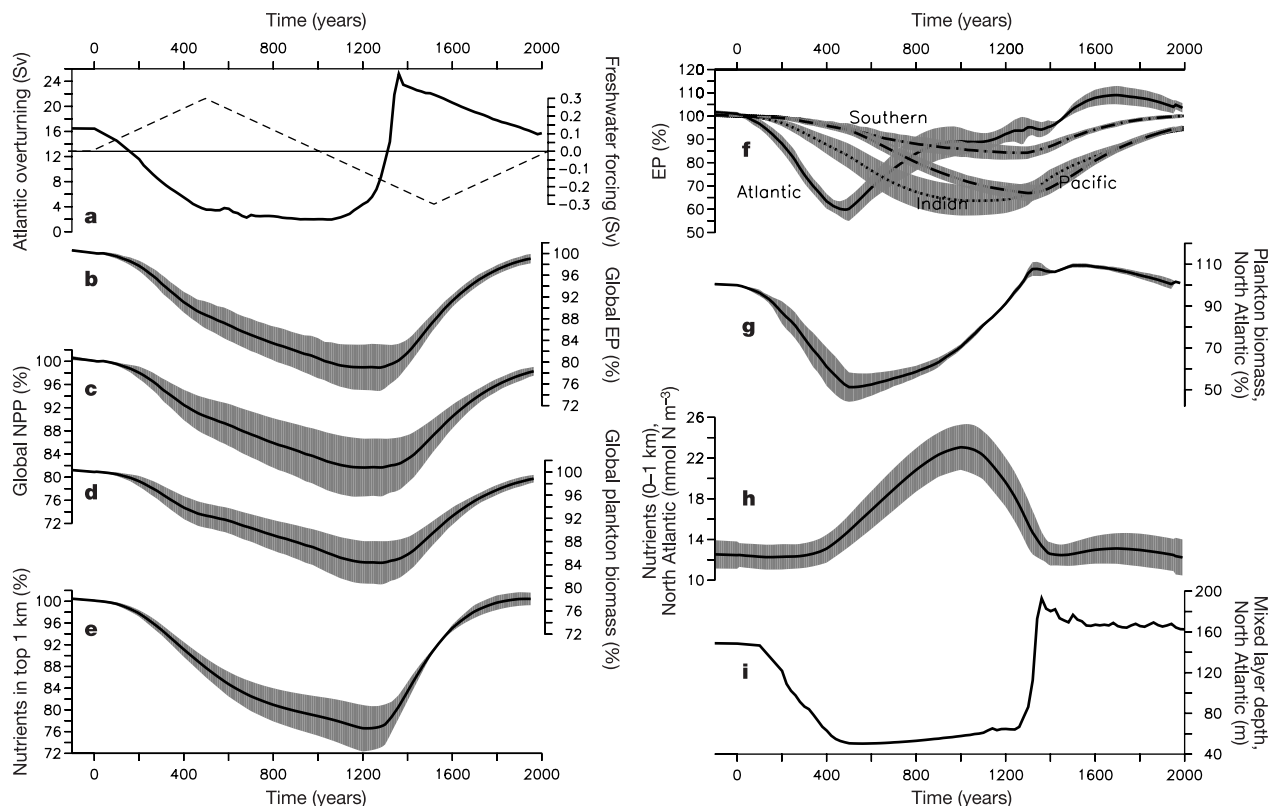


Figure 1 Time series from the model simulations. Left panels show freshwater forcing in the North Atlantic (dashed line, right scale) and AMO circulation (solid line, left scale) (a), globally integrated export production, EP (b), net primary production, NPP (c), plankton (phytoplankton plus zooplankton) stocks (d) and nutrient concentrations in the top 1 km (e). Right panels show changes in integrated export production (f) in the individual ocean basins (Atlantic, solid; Pacific, dashed; Indian, dotted; Southern Ocean, dash-dotted) as well as integrated biomass (g), average nitrate concentrations (h) and average mixed layer

depth (calculated as depth of $\Delta\sigma_\theta = 0.25$ using annual mean densities) in the North Atlantic from 35° N to 70° N (i; rectangle in Fig. 2). Panels b–g give percentage changes with respect to year 0. The grey bars enclosing the ensemble mean (solid lines) denote the standard deviation of the different ecosystem models used and represent the uncertainties associated with the formulation and parameterizations of the ecosystem models. The evolution of the physical variables does not vary between the different experiments, hence panels a and i do not contain shadings.

nutrient budget confirmed (not shown). In the North Atlantic, increasing subsurface nutrient levels lead to a gradual recovery of plankton stocks after year 500 (Fig. 1g). Around year 1,200, that is, 700 yr after the collapse, North Atlantic plankton stocks re-attain their initial size. Note that this occurs before the resumption of the circulation around year 1,340.

The simulated evolution of productivity in the North Atlantic (Fig. 4a) is consistent with recent reconstructions^{10–12}. A qualitative productivity record of high temporal resolution from the Reykjanes ridge (60° N, 30° W), based on planktonic and benthic foraminifera species composition and accumulation rates, indicates low productivity during stadials and high productivity during interstadials¹¹. This is in agreement with the simulations, which show stadial productivity reduced by about 40%. Consistently, a record of benthic species composition from the northeast Atlantic (50° N, 20° W) also indicates low productivity during Heinrich events¹². A quantitative reconstruction of productivity from the northeast Atlantic¹⁰ (53° N, 19° W) shows about 50% decrease during cold and fresh episodes with respect to warm and saline periods, consistent with the simulations.

A fast reduction of productivity and planktonic biomass is also simulated in the equatorial and South Atlantic (Figs 2d, 4a). In the western equatorial Atlantic (61° W, 12° N), productivity decreases

by 25–50%, consistent with a high-resolution record of dinoflagellate cyst and organic carbon accumulation at this site. The reduced productivity during the stadial Dansgaard–Oeschger phases¹³ has been attributed to variations in the Orinoco River nutrient discharge and coastal upwelling strength. These processes are not included in the simulation, which shows a decrease in response to thermohaline driven changes in ocean circulation only. In the eastern South Atlantic, productivity drops dramatically during the first 500 yr of the simulations and remains low throughout the stadial. Reversal of the upper ocean flow direction (Fig. 3c, e) from mainly northward at year 0 to southward after year 500 cuts off the South Atlantic from a supply of high-nutrient waters of southern origin. Additionally, reduced upwelling decreases nutrients in the southeast Atlantic.

Productivity also declines in the Arabian Sea and the eastern tropical Indian Ocean, as well as in the Southern Ocean. In the Arabian Sea at 66° E, 23° N, for example, productivity decreases by about 40% until year 1,200 (Fig. 4b). This is consistent with recent alkenone-based reconstructions from this site¹⁴ suggesting higher productivity during interstadials and lower productivity during stadials. However, in the simulation this is caused by purely thermohaline-driven changes in upwelling and not through wind-driven upwelling as suggested previously¹⁴.

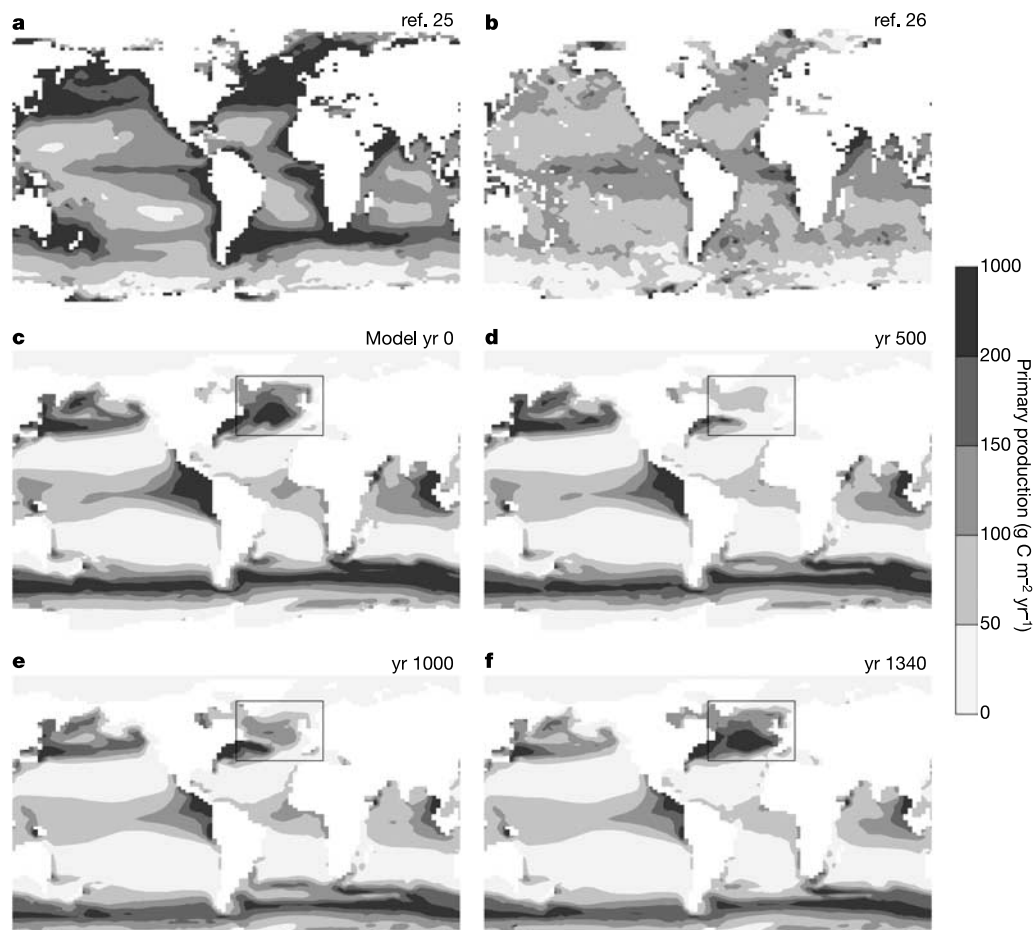


Figure 2 Annual net primary productivity from observations and the ensemble mean of the simulations. Differences between the two estimates based on satellite observations^{25,26} (a, b) represent mainly uncertainties associated with algorithms used to infer vertically integrated productivity from ocean colour (chlorophyll) observations. The ensemble mean during the initial state of an operating AMO at year 0 (c) is compared with

the satellite estimates in the Methods section. Snapshots during the simulations show the collapsed AMO at year 500 at minimum productivity in the North Atlantic (d), the state at year 1,000 during minimum global productivity (e) and the situation during the rapid recovery at year 1,340 (f). The squares denote the area in the North Atlantic for which averaged time series are shown in Fig. 1.

After entering the stadial state at year 500, the ecosystem also declines in the North Pacific, the eastern tropical Pacific and in the Southern Ocean owing to gradual depletion of surface nutrients there (Figs 3, 4). High-resolution records from the northeast Pacific demonstrate anoxic conditions during interstadials and better oxygenation during stadials^{15,16}. This has been explained by changes in intermediate water mass ventilation¹⁵ and/or surface water productivity¹⁶. In the simulation, productivity at 120° W, 34° N decreases by up to 50% from interstadial to stadial levels. Lower productivity leads to slower rates of remineralization of organic matter, and therefore to reduced consumption of oxygen at depth. Thus, the simulation is consistent with the record, and supports the hypothesis that reduced productivity was at least partly responsible for the higher oxygen concentrations during the cold events. However, contrary to speculations that changes in ENSO (El Niño/Southern Oscillation) were responsible for the stadial productivity minimum¹⁶, in the simulation it is solely caused by thermohaline-driven reduction in upwelling.

The AMO spindown does not decrease productivity everywhere.

Along the eastern boundary current of the North Atlantic subtropical gyre, for example, productivity increases during the first 750 yr of the simulation (Figs 2, 4). At 20° W, 26° N (the location of sediment core 15637) and at 20° W, 40° N (along the Iberian margin), productivity increases by up to a factor of five during the stadial. Palaeoproductivity records from these sites^{17,18} indicate increased productivity by a factor of two or more during cold events, consistent with the model results. Analysis of the simulation indicates that increasing subsurface nutrient concentrations and stronger upwelling fertilize the ocean along the Iberian and North African coasts during the stadials.

Taken together (Fig. 4), the available proxy record corroborates the model results. However, a more stringent and quantitative test is at present hampered by the low temporal and spatial resolution of existing productivity records. Particularly desirable would be additional records from the South(east) Atlantic, the Southern Ocean and from the tropical, south and northwest Pacific.

This Letter has examined the influence of circulation changes on the marine ecosystem. In the case of enhanced future global

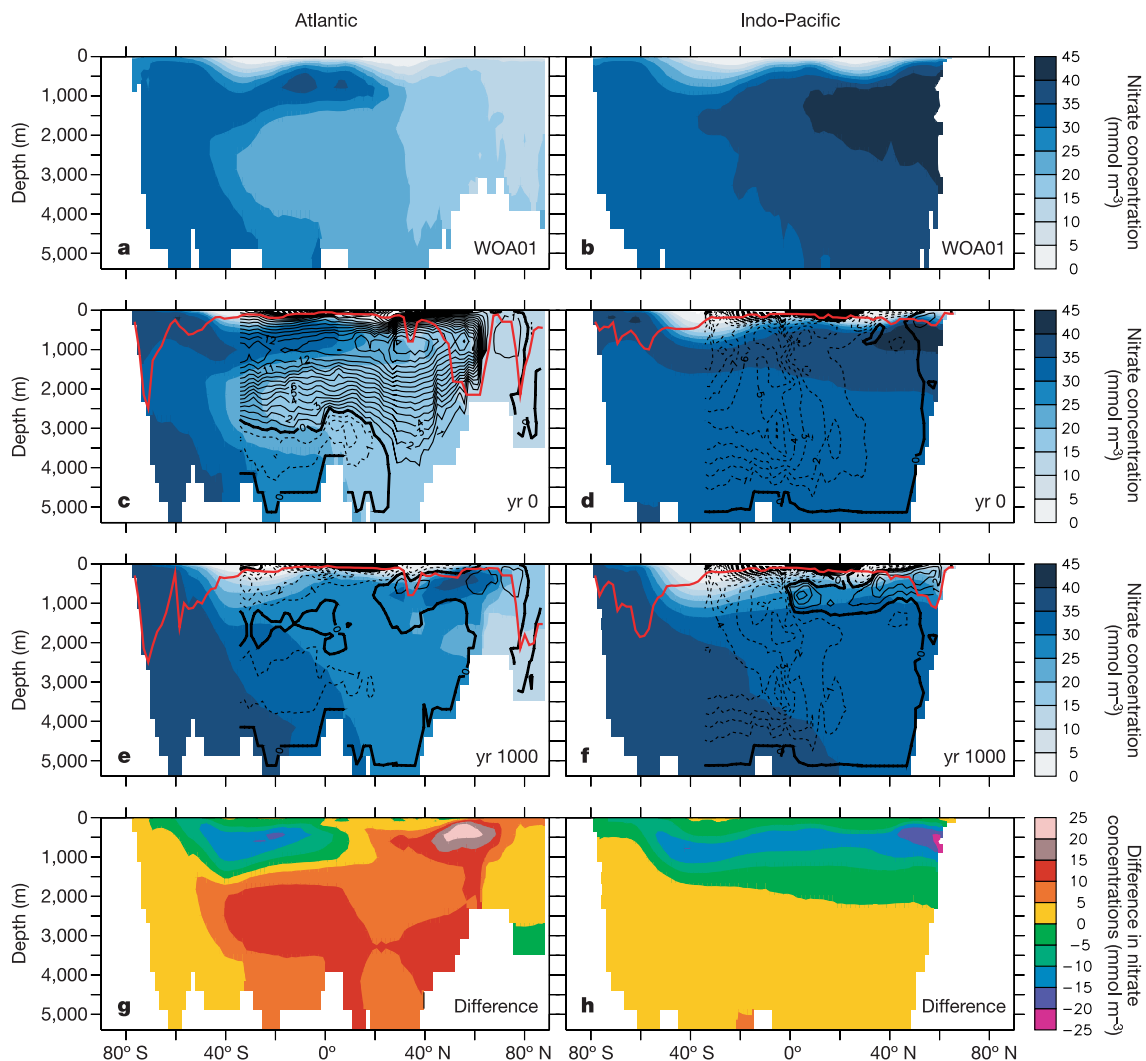


Figure 3 Latitude-depth sections of zonally averaged nitrate concentrations in mmol m^{-3} . Modern observations²⁸ (**a**, **b**) and model results from year 0 (**c**, **d**), at year 1,000 (**e**, **f**) and the difference between year 1,000 and year 0 (**g**, **h**) in the Atlantic (left) and Indo-Pacific (right). Black contour lines give the zonally integrated eulerian circulation in sverdrups ($1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$), and red lines denote the zonal and annual maximum monthly mixed layer depth (calculated as depth of $\Delta\sigma_\theta = 0.1 \text{ kg m}^{-3}$). In the case of an

active overturning (**a**, **c**), the North Atlantic is nutrient depleted owing to extraction of nutrients along the northward path of the upper branch of the circulation. Low-nutrient North Atlantic Deep Water (NADW) then flows southward at around 2,000 m depth. After the AMO stops, nutrient levels increase in the North Atlantic (**e**, **g**) and a subsurface nutrient maximum develops much like in today's North Pacific (**b**, **d**), while concentrations decrease in the top 1 km of the South Atlantic, Southern and Indo-Pacific oceans (**e**–**h**).

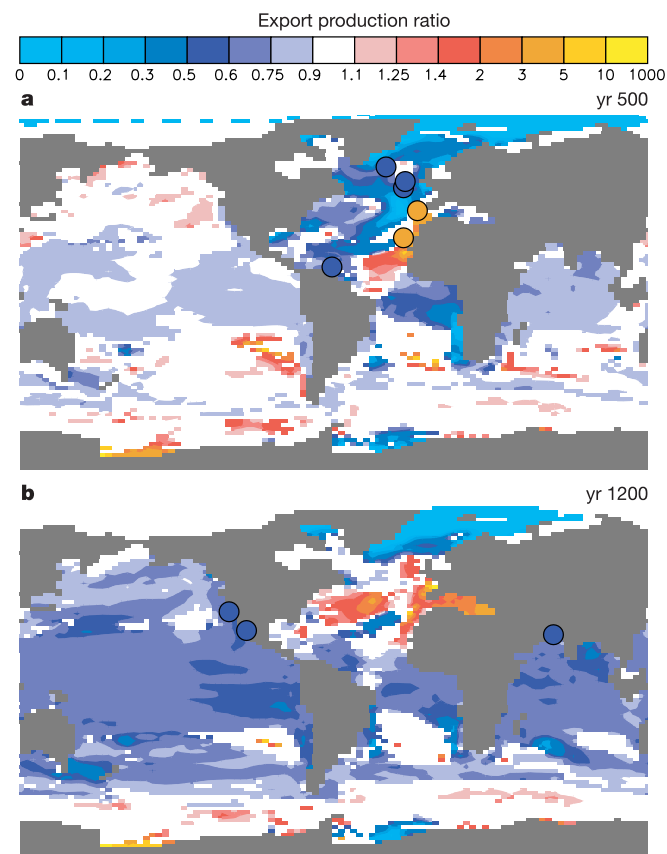


Figure 4 Comparison of simulated changes in export production with palaeoproductivity proxy records. Coloured circles show the proxy records. Colour contours give the model ensemble mean ratio of export production during the early stadial stage (**a**; yr 500) and the late stadial stage (**b**; yr 1,200) with respect to interstadial (model year 0) values. A value of 0.3, for example, indicates export production decreased by 70%. Only statistically significant values are plotted, in the sense that the mean $\pm$ the standard deviation is not allowed to include one (no change). Blue coloured circles denote lower productivity during stadials, and orange coloured circles denote increased productivity as inferred from the proxy records discussed in the main text. Note that only qualitative changes are shown for the palaeorecords, as most reconstructions give no quantitative estimates. As the largest productivity changes are simulated in the North Atlantic during the early stadial stage (see Fig. 1) while in the Indian and Pacific oceans the largest changes occur during the late stadial stage, the Atlantic proxy records are shown in the upper panel^{10–13,17,18} and the Indian and Pacific records in the lower panel^{14–16}.

warming, additional effects such as increased stratification due to upper ocean warming will affect marine productivity¹⁹. Moreover, in the simulations presented above, the AMO spindown is relatively slow (about 500 yr, see Fig. 1). More rapid reductions as projected by some models^{5–7} would presumably lead to an even stronger decline in biomass, particularly in the North Atlantic. From Fig. 1 it appears that even a partial weakening of the overturning causes a substantial reduction of productivity. The results presented above have important implications for the assessment of future greenhouse gas emission scenarios. A massive decline of plankton stocks could have catastrophic effects on fisheries and human food supply in the affected regions. Hence, emission pathways that lead to fast and large increases of future CO₂ including the risk of a collapse or substantial reduction of the AMO⁶ should be avoided through early measures for emission reductions. □

Methods

Climate model

For the simulations presented in this Letter, I use the University of Victoria Earth System Climate Model²⁰. This is a global model consisting of a fully nonlinear primitive equations

three-dimensional ocean circulation model using a scheme of tidal mixing²¹ (which leads to low values of vertical diffusion of $0.2\text{--}0.3\text{ cm}^2\text{ s}^{-1}$ within the pycnocline), a dynamic thermodynamic state-of-the-art sea ice component and a single level energy moisture balance model of the atmosphere. A prescribed seasonal cycle of present-day near surface atmospheric wind velocities is used in the advection of specific humidity, for the momentum transfer to the surface ocean and sea ice and in the calculation of surface sensible and latent heat fluxes. As such, the atmospheric hydrological cycle is interactive and considers thermodynamic effects (Clausius–Clapeyron) but neglects dynamic effects (for example, changes in the Hadley circulation). The model response might therefore be different from a model with a more complete, dynamical atmospheric component. The simulations presented above use a preindustrial (interglacial) background climate (for example, present-day sea level, atmospheric CO₂ concentration of 280 p.p.m.v.). However, an equivalent simulation under glacial conditions confirmed the results presented.

Ecosystem models and simulations

The models of the marine ecosystem dynamics, solving prognostic equations for nutrients, phytoplankton, zooplankton, detritus and dissolved organic matter, based on nitrate as the limiting nutrient, have been interactively coupled to the ocean circulation component of the model²². In order to consider uncertainties in the formulation of the ecosystem models, identical (ensemble) simulations have been performed with four different models. The four ensemble members represent two versions from ref. 22 (termed ‘REF’ and ‘tidal $K_b = 0.2$, $\sigma_{DOM} = 0.15$ $\mu_{DOM} = 0.17$ ’ in Table 1 of ref. 22), the formulation of ref. 23 and the model of ref. 24. The difference between the first two is the inclusion of dissolved organic matter cycling, and the last two models use fundamentally different formulations and parameterizations. All four models produce similarly good agreement with global nutrient (Fig. 3) and chlorophyll (phytoplankton) observations. The model of ref. 24 uses a parameterization of fast microbial nutrient recycling through ammonium which increases productivity mainly at low latitudes. This leads to global primary productivity of 48.6 Gt C yr^{-1} , which is in better agreement with observations ($36\text{--}50\text{ Gt C yr}^{-1}$)^{25,26} than the values of $20\text{--}30\text{ Gt C yr}^{-1}$ of the other models. The ensemble mean primary productivity for modern conditions (Fig. 2c) is broadly consistent with satellite estimates within their uncertainties (Fig. 2a, b). Areas of high productivity are the mid-latitudes of the Northern Hemisphere, the equatorial upwelling areas, the Southern Ocean and the zonal boundary currents.

Systematic biases occur in the centres of the subtropical gyres, where most models are too oligotrophic. The Southern Ocean is too eutrophic in most models, and the belt of maximum productivity is located too far south. Despite the large differences between the individual models, certain aspects (such as nitrogen fixation and denitrification, as well as limitation by nutrients other than nitrate) are not taken into account in either model. It is therefore possible, that for example, the neglect of iron limitation biases the results in iron-limited regions. But because the vertical profile of iron in the ocean is similar to that of the other nutrients (low values in the euphotic zone and high concentrations at depths), and because it has been estimated that only 25% of the iron supply to the phytoplankton comes from atmospheric deposition whereas 75% is supplied through upwelling²⁷, it can be assumed that reduced upwelling also leads to a reduced iron supply. This suggests that missing iron limitation does not qualitatively affect the results. Generally, the reliability of the results presented above will be better in nitrate-limited regions, such as the North Atlantic, and probably less in regions that are thought to be limited by other nutrients (like the iron-limited Southern Ocean and the equatorial Pacific). Note also that the sensitivity of the AMO to perturbations might be strongly model dependent, as indicated by the large scatter in future projections⁷.

Received 6 August 2004; accepted 7 February 2005; doi:10.1038/nature03476.

- Clark, P. U., Pisias, N. G., Stocker, T. F. & Weaver, A. J. The role of the thermohaline circulation in abrupt climate change. *Nature* **415**, 863–869 (2002).
- Sarnthein, M. et al. in *The Northern North Atlantic: A Changing Environment* (eds Schäfer, P., Ritzau, W., Schlüter, M. & Thiede, J.) 45–66 (Springer, Berlin, 2001).
- Dokken, T. M. & Jansen, E. Rapid changes in the mechanism of ocean convection during the last glacial period. *Nature* **401**, 458–461 (1999).
- Schmittner, A., Saenko, O. A. & Weaver, A. J. Coupling of the hemispheres in observations and simulations of glacial climate change. *Quat. Sci. Rev.* **22**, 659–671 (2003).
- Manabe, S. & Stouffer, R. J. Century-scale effects of increased atmospheric CO₂ on the ocean-atmosphere system. *Nature* **364**, 215–218 (1993).
- Stocker, T. F. & Schmittner, A. Influence of CO₂ emission rates on the stability of the thermohaline circulation. *Nature* **388**, 862–865 (1997).
- Cubasch, U. et al. in *Climate Change 2001: The Scientific Basis* (ed. Houghton, J. T.) 525–582 (Cambridge Univ. Press, Cambridge, 2001).
- Willamowski, C. & Zahn, R. Upper ocean circulation in the glacial North Atlantic from benthic foraminiferal isotope and trace element fingerprinting. *Paleoceanography* **15**, 515–527 (2000).
- Sarmiento, J. L., Gruber, N., Brzezinski, M. A. & Dunne, J. P. High-latitude controls of thermocline nutrients and low latitude biological productivity. *Nature* **427**, 56–60 (2003).
- Weinelt, M., Rosell-Mele, A., Pflaumann, U., Sarnthein, M. & Kiefer, T. Zur Rolle der Produktivität im Nordatlantik bei abrupten Klimaänderungen in den letzten 8000 Jahren. *Z. Dt. Geol. Ges.* **154**, 47–66 (2003).
- Rasmussen, T. L., Thomson, E., Troelstra, S. R., Kuijpers, A. & Prins, M. A. Millennial-scale glacial variability versus Holocene stability: changes in planktic and benthic foraminiferal faunas and ocean circulation in the North Atlantic during the last 60 000 years. *Mar. Micropaleont.* **47**, 143–176 (2002).
- Thomas, E., Booth, L., Maslin, M. & Shackleton, N. J. Northeastern Atlantic benthic foraminifera during the last 45,000 years: Changes in productivity seen from the bottom up. *Paleoceanography* **10**, 545–562 (1995).
- Vink, A. et al. Shifts in the position of the North Equatorial Current and rapid productivity changes in the western Tropical Atlantic during the last glacial. *Paleoceanography* **16**, 479–490 (2001).

14. Schulte, S. & Müller, P. J. Variations of sea surface temperature and primary productivity during Heinrich and Dansgaard-Oeschger events in the northeastern Arabian Sea. *Geo-Mar. Lett.* **21**, 168–175 (2001).
15. Hendy, I. L. & Kennett, J. P. Tropical forcing of North Pacific intermediate water distribution during Late Quaternary rapid climate change? *Quat. Sci. Rev.* **22**, 673–689 (2003).
16. Ortiz, J. D. *et al.* Enhanced productivity off western North America during the warm climate intervals of the past 52 kyr. *Geology* **32**, 521–524 (2004).
17. Kiefer, T. *Productivity and Temperatures in the Subtropical North Atlantic: Cyclic and Abrupt Changes During the Late Quaternary* (Rep. No. 90, Geol.-Paläont. Inst. Univ. Kiel, 1998).
18. Lebreiro, S. M., Moreno, J. C., Abrantes, F. F. & Pflaumann, U. Productivity and paleoceanographic implications on the Tore Seamount (Iberian Margin) during the last 225 kyr: Foraminiferal evidence. *Paleoceanography* **12**, 718–727 (1997).
19. Bopp, L. *et al.* Potential impact of climate change on marine export production. *Glob. Biogeochem. Cycles* **15**, 81–100 (2001).
20. Weaver, A. J. *et al.* The UVic Earth system climate model: Model description, climatology and applications to past, present and future climates. *Atmosphere-Ocean* **4**, 361–428 (2001).
21. Simmons, H. L., Jayne, S. R., St Laurent, L. C. & Weaver, A. J. Tidally driven mixing in a numerical model of the ocean general circulation. *Ocean Model.* **6**, 245–263 (2004).
22. Schmittner, A., Oschlies, A., Giraud, X., Eby, M. & Simmons, H. L. A global model of the marine ecosystem for long term simulations: sensitivity to ocean mixing, buoyancy forcing, particle sinking and dissolved organic matter cycling. *Glob. Biogeochem. Cycles* (in the press).
23. Six, K. D. & Maier-Reimer, E. Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model. *Glob. Biogeochem. Cycles* **10**, 559–583 (1996).
24. Schartau, M. & Oschlies, A. Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part I—Method and parameter estimates. *J. Mar. Res.* **61**, 765–793 (2003).
25. Gregg, W. W., Conkright, M. E., Ginoux, P., O'Reilly, J. E. & Casey, N. W. Ocean primary production and climate: Global decadal changes. *Geophys. Res. Lett.* **30**, 1809, doi:10.1029/2003GL018889 (2003).
26. Antoine, D., André, J.-M. & Morel, A. Oceanic primary production 2. Estimation at global scale from satellite (coastal zone color scanner) chlorophyll. *Glob. Biogeochem. Cycles* **10**, 57–69 (1996).
27. Aumont, O., Maier-Reimer, E., Blain, S. & Monfray, P. An ecosystem model of the global ocean including Fe, Si, P colimitations. *Glob. Biogeochem. Cycles* **17**, 1060, doi:10.1029/2001GB001745 (2003).
28. Conkright, M. E. *et al.* *World Ocean Atlas 2001, Objective Analysis, Data Statistics, and Figures* (CD-ROM Documentation, National Oceanographic Data Center, Silver Spring, Maryland, 2002).

Acknowledgements Productivity data sets were provided by D. Antoine and W. Gregg. Discussions with A. Oschlies, M. Sarnthein, M. Weinelt and H. Kinkel were appreciated. This research was supported as part of the research unit (Forschergruppe 451) on ocean gateways by the Deutsche Forschungsgemeinschaft (DFG).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.S. (aschmittner@coas.oregonstate.edu).

Structural and temporal requirements for geomagnetic field reversal deduced from lava flows

Brad S. Singer¹, Kenneth A. Hoffman², Robert S. Coe³, Laurie L. Brown⁴, Brian R. Jicha¹, Malcolm S. Pringle⁵ & Annick Chauvin⁶

¹Department of Geology and Geophysics, University of Wisconsin, Madison, Wisconsin 53706, USA

²Physics Department, California Polytechnic State University, San Luis Obispo, California 93407, USA

³Earth Sciences Department, University of California, Santa Cruz, California 95064, USA

⁴Department of Geosciences, University of Massachusetts, Amherst, Massachusetts 01003, USA

⁵Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

⁶Geosciences Rennes UMR-CNRS, Campus Beaulieu, Bâtiment 15, 35042 Rennes Cedex, France

Reversals of the Earth's magnetic field reflect changes in the geodynamo—flow within the outer core—that generates the field. Constraining core processes or mantle properties that induce or modulate reversals requires knowing the timing and morphology of field changes that precede and accompany these reversals^{1–4}. But the short duration of transitional field states

and fragmentary nature of even the best palaeomagnetic records make it difficult to provide a timeline for the reversal process^{1,5}. ⁴⁰Ar/³⁹Ar dating of lavas on Tahiti, long thought to record the primary part of the most recent 'Matuyama–Brunhes' reversal, gives an age of 795 ± 7 kyr, indistinguishable from that of lavas in Chile and La Palma that record a transition in the Earth's magnetic field, but older than the accepted age for the reversal. Only the 'transitional' lavas on Maui and one from La Palma (dated at 776 ± 2 kyr), agree with the astronomical age for the reversal. Here we propose that the older lavas record the onset of a geodynamo process, which only on occasion would result in polarity change. This initial instability, associated with the first of two decreases in field intensity, began ~18 kyr before the actual polarity switch. These data support the claim⁶ that complete reversals require a significant period for magnetic flux to escape from the solid inner core and sufficiently weaken its stabilizing effect⁷.

Most reversal records come from quasi-continuously deposited sediments^{1,8–9} and many indicate that reversals occur when field intensity has diminished (see, for example, refs 10, 11). Establishing the rate of sediment accumulation may reveal a duration for the reversal process, but accumulation rates are difficult to quantify¹. One approach is to tune oxygen isotope variations orbitally down the section, estimate the astronomical ages of successive magnetic reversals¹² and interpolate the accumulation rate. With the use of

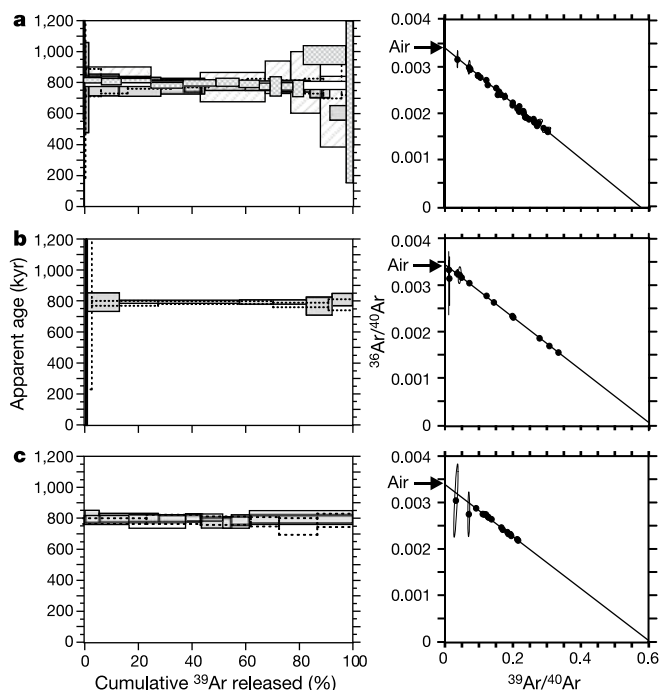


Figure 1 ⁴⁰Ar/³⁹Ar age spectra and isochrons from lavas TT, R1T and R1V in Punaruu Valley, Tahiti. The largely concordant spectra (left) and well-defined isochrons (right) indicate that extraneous argon is not a problem. Isochrons of several experiments on each lava were normalized to a common neutron fluence parameter *J* for illustrative purposes only. The weighted mean of the individual isochrons from several subsamples gives the best age (±2σ) for each lava. **a**, TT groundmass. The plateau is at 789.7 ± 5.2 kyr; the results are from five separate experiments. The isochron age is 798.0 ± 11.0 kyr, ⁴⁰Ar/³⁶Ar = 293.9 ± 3.6, MSWD = 0.15 and *n* = 32 of 39. **b**, R1V groundmass. The plateau is at 789.4 ± 6.5 kyr; the results are from two separate experiments. The isochron age is 791.9 ± 9.3 kyr, ⁴⁰Ar/³⁶Ar = 294.8 ± 1.9, MSWD = 0.04 and *n* = 13 of 13. **c**, R1T groundmass. The plateau is at 792.1 ± 7.6 kyr; the results are from three separate experiments. The isochron age is 798.0 ± 23.0 kyr, ⁴⁰Ar/³⁶Ar = 294.6 ± 2.9, MSWD = 0.70 and *n* = 18 of 18.

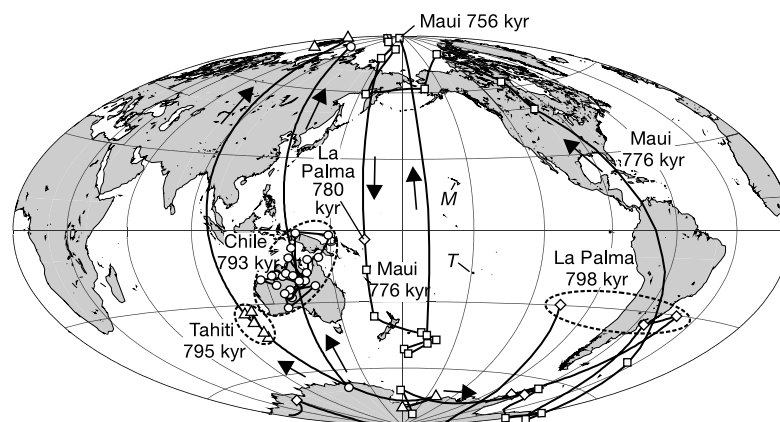


Figure 2 Virtual Geomagnetic Poles (VGPs) and paths of four lava sequences that record transitional directions between 795 and 775 kyr ago. Mean ages of each lava sequence are from Table 1 in thousands of years. The youngest sequence at Maui begins with VGPs in Antarctica that progress through the Americas, sweep clockwise through Alaska

and then move down to a cluster southeast of New Zealand; it is capped by a long series of normally magnetized flows, of which the lowest is dated at 756 kyr ago. M and T indicate the locations of Maui and Tahiti, respectively.

this method on several sediment cores, the average duration of the last four reversals, including the most recent Matuyama–Brunhes (M–B) reversal, is $6,992 \pm 1,089$ yr (2σ)¹³. However, the duration is dependent on the latitude of the observation site, increasing from low latitudes polewards from about 2 to 10 kyr¹³.

Thermoremanent magnetization of lava flows also records reversals; however, each lava in a sequence of flows provides only an instantaneous ‘spot’ recording of the field, so no single sequence is able to capture a reversal in its entirety^{1,4,14}. Notwithstanding this, an advantage is that lavas can be dated precisely using the $^{40}\text{Ar}/^{39}\text{Ar}$ variant of the K–Ar geochronometer^{15–18}. Our approach to resolving field changes attending a reversal is to examine the palaeomagnetic directions and $^{40}\text{Ar}/^{39}\text{Ar}$ chronology of four lava sequences that presumably erupted during the M–B reversal.

Three of these sequences met criteria for incorporation into a

global database for the M–B reversal¹⁹: Punaruu Valley, Tahiti^{20,21}; Tataru–San Pedro, Chile^{18,21–22}; and Haleakala, Maui^{17,21}. The fourth sequence on La Palma, Canary Islands, was characterized recently¹⁶. On the basis of the ages of eight lavas from these four sequences, used originally to argue that the duration of the M–B reversal was at least 12 kyr (ref. 21), our goal was to expand the data set and test the hypothesis that the duration of the reversal exceeds the ~7,000-yr average inferred from sediments¹³. We present new $^{40}\text{Ar}/^{39}\text{Ar}$ ages of transitionally magnetized lavas from Tahiti alongside ages from the other three lava recordings and palaeointensity records from marine sediment, which together illustrate that the change in polarity may be a late stage of the reversal process.

The same $^{40}\text{Ar}/^{39}\text{Ar}$ incremental heating method^{15–18} and inter-calibrated standard values²³ employed to date 20 transitionally magnetized lavas on Maui, Chile and La Palma were used to date

Table 1 $^{40}\text{Ar}/^{39}\text{Ar}$ incremental heating ages of transitionally magnetized lavas

Location	Flow unit	Ref.	VGP		No. of experiments			Concordant increments	% of total ^{39}Ar	Isochron calculation		
			Lat.	Long.	UW	GE	SU			MSWD	$^{40}\text{Ar}/^{36}\text{Ar}_i \pm 2\sigma$	Age (kyr) $\pm 2\sigma$
Haleakala	59	17	–44.2	190.8	3	1	2	50 of 59	94.7	0.96	296.5 ± 1.3	774.2 ± 3.6
	58	17	–47.9	191.3	2	2	1	34 of 45	79.5	1.16	295.7 ± 1.1	778.3 ± 5.2
	52	17	–38.2	163.3			1	8 of 13	78.0	11.73	293.4 ± 7.2	782.4 ± 26.4
	50	17	75.7	142.8			1	9 of 14	82.1	0.34	296.8 ± 1.5	778.7 ± 6.9
	45	17	77.5	222.8			1	9 of 14	85.1	1.38	293.9 ± 0.8	785.1 ± 8.0
	37	17	–67.5	180.9	5		1	47 of 61	87.5	1.30	295.3 ± 0.8	773.0 ± 3.0
Weighted mean age for six lavas from Haleakala												
Chile	QW11-17	18	–9.8	136.0	3			17 of 18	96.3	0.20	295.4 ± 0.8	801.0 ± 12.0
	QW11-16	18	–24.8	137.8	2			17 of 22	92.6	1.02	294.2 ± 1.8	791.3 ± 7.9
	QW11-11	18	–26.1	143.7	1			9 of 13	82.7	1.02	295.5 ± 4.5	792.9 ± 17.1
	QW11-5	18	–25.7	126.3	3			20 of 24	88.8	1.44	293.6 ± 3.2	793.8 ± 9.8
	QW11-3	18	–21.8	130.3	2			12 of 23	68.0	0.68	292.6 ± 4.1	795.0 ± 21.0
	QW10-10	18	–18.9	147.6		1		8 of 11	76.9	0.42	296.7 ± 5.4	790.1 ± 12.6
	QW10-8	18	–18.7	156.4		3		19 of 26	80.6	0.76	294.4 ± 2.7	790.8 ± 6.7
	QW10-5	18	–14.4	145.5			2	11 of 13	86.3	1.09	297.5 ± 3.8	790.3 ± 8.6
	QW10-3	18	–3.4	138.2	2	3		44 of 57	77.5	2.00	292.2 ± 2.8	790.2 ± 6.1
Weighted mean age for nine Chilean lavas												
Tahiti	TT		–34.0	104.6	3	2		32 of 39	83.5	0.15	293.9 ± 3.6	798.0 ± 11.0
	R1V		–33.1	108.6	2			13 of 13	100.0	0.04	294.8 ± 1.9	791.9 ± 9.3
	R1T		–41.9	107.9	3			18 of 18	100.0	0.70	294.6 ± 2.9	798.0 ± 23.0
Weighted mean age for three Tahitian lavas												
La Palma	TN-16	16	–5.4	161.6	5			43 of 59	81.2	1.41	296.3 ± 1.3	780.3 ± 10.3
	TN-17	16	–31.1	250.9	5			34 of 46	82.9	1.40	293.9 ± 1.5	803.3 ± 9.3
	TN-18	16	–70.3	52.0	5			56 of 56	100.0	1.01	294.7 ± 1.7	796.2 ± 10.3
	TN-20b	16	–31.6	313.3	4			39 of 50	91.4	1.81	294.0 ± 2.5	791.2 ± 18.9
	CB202a	21	–48.0	301.0		2		15 of 23	81.0	1.29	299.8 ± 3.8	786.2 ± 27.2
Weighted mean age for four lavas from La Palma												
											0.70	798.4 ± 6.3

All ages are relative to a common 28.34-Myr-old Taylor Creek sanidine standard²³, reported with analytical uncertainty only. Systematic uncertainties in, for example, the ^{40}K decay constant, need not be considered when comparing between these ages. Laboratories: UW, Wisconsin; GE, Geneva; SU, Scotland.

three of the five transitional lavas in the Punaruu section. To improve precision, ages and 2σ analytical uncertainties are calculated as the inverse-variance weighted mean obtained from one to six subsamples of each lava^{15–18}. Eight new experiments yielded mainly concordant age spectra, well-defined isochrons of 798.0 ± 23.0 , 791.9 ± 9.3 and 798.0 ± 11.0 kyr, and show no evidence for inherited or excess argon (Fig. 1; see Supplementary Information). The Punaruu sequence describes a reverse–transitional–normal (R–T–N) directional path with clustering of five virtual geomagnetic poles (VGPs) southwest of Australia²⁰ (Fig. 2). The similar VGPs and ages give reason^{15–18} to use the weighted mean of the three isochrons of 794.8 ± 6.8 kyr as the most precise estimate of time since the transitional field was recorded (Table 1).

At La Palma, seven basaltic flows record a R–T–R–T–N sequence of VGPs¹⁶ (Fig. 2). The ages of the three lowermost transitional lavas, with VGPs over southern South America, are indistinguishable from one another and give a weighted mean of 798.4 ± 6.3 kyr that is also not different from that of the three Tahitian lavas (Table 1). The stratigraphically highest transitional lava TN-16 at La Palma yields an age of 780.3 ± 10.3 kyr, which is significantly younger than the mean of the lower three lavas in the section, and younger than the Tahitian lavas (Table 1). Moreover, lava TN-16 has a VGP northeast of Australia that differs from the underlying flows¹⁶ (Fig. 2).

At Tatara-San Pedro volcano 30 andesitic lava flows exhibit R–T–N directions with a tight clustering of VGPs over northern Australia¹⁸ (Fig. 2). Earlier K–Ar²² and $^{40}\text{Ar}/^{39}\text{Ar}$ ages²¹ implied that these lavas record the M–B reversal. Nine of these lavas yield a weighted mean age of 791.7 ± 3.0 kyr (ref. 18) identical to those obtained at La Palma and Tahiti (Table 1).

A section of basaltic lavas at Haleakala volcano includes 24 with VGP latitudes that cluster into three groups, including six between -80° and -45° at the base, seven between $+40^\circ$ and 75° in the middle, followed by eight between -35° and -50° near the top¹⁷ (Fig. 2). $^{40}\text{Ar}/^{39}\text{Ar}$ ages of six flows are indistinguishable from one another and give a weighted mean of 775.6 ± 1.9 kyr (Table 1), indicating that this sequence corresponds to the M–B reversal¹⁷. The mean age of the Maui sequence is significantly younger than that of the other three sections, with one exception: the uppermost TN-16 lava from La Palma (Table 1).

The non-gaussian distribution of these lava ages reveals itself first by the excessive mean square of weighted deviates (MSWD) of 6.7 associated with the weighted mean age of 781.9 ± 1.5 kyr when all 23 isochrons in Table 1 are combined, and second, by the bimodal structure of the probability density distribution (Fig. 3). The older peak is defined by the three lowermost lavas at La Palma, plus the three Tahitian and nine Chilean ages, which are indistinguishable from one another and yield a weighted mean of 793.2 ± 2.5 kyr (MSWD = 0.7). In contrast, the weighted mean age of the six Haleakala lavas plus the TN-16 lava from La Palma is 775.5 ± 1.9 kyr (MSWD = 2.0); the difference between the younger and older groups of lavas is 17.7 ± 3.1 kyr.

Palaeointensity measurements from ODP site 983, where sedimentation rates of 13 cm kyr^{-1} maximize temporal resolution, show a prominent dip in intensity before the change in VGP and a concomitant drop in intensity that reflects the M–B reversal¹¹. The astronomical ages of the M–B reversal and the preceding dip in intensity determined at site 983 are 775 and 793 kyr, respectively¹¹. Evidently, this early dip in palaeointensity is a global feature of the magnetic field that, on the basis of astrochronological dating of a dozen Atlantic and Pacific cores^{10,24–25}, occurred ~ 15 kyr before the M–B reversal (Fig. 3).

The $^{40}\text{Ar}/^{39}\text{Ar}$ age of 776 ± 2 kyr for the Maui/upper La Palma lavas is identical to the astrochronological age of the M–B reversal of 778 ± 4 kyr globally²⁵, and 775 kyr at site 983 (ref. 11). Further, the older $^{40}\text{Ar}/^{39}\text{Ar}$ age of 793 ± 3 kyr for the lavas in Tahiti, Chile and

La Palma is identical to the astrochronological age of the precursor event both globally²⁵ and at ODP site 983 (ref. 11). Clement's finding¹³ that reversal duration is dependent on site latitude might suggest that even precise age determinations on transitionally magnetized lavas from distinct sites, each unable to continuously record field behaviour throughout a reversal, could result in spatial–temporal misinterpretation of the dynamo process. However, given both the synchronism between marine sediments^{11,25} and transitional lavas, and the proximity of Maui and Tahiti—hotspot islands only 38° apart (Fig. 2)—this is clearly not so in the present case. We therefore conclude that the transitional lavas from Tahiti, Chile and La Palma do not record the final directional reversal of the M–B transition, but instead provide a fragmentary snapshot of directions associated with the weakening of the field that preceded the reversal by 18 kyr.

We now argue that this extended period of unstable field behaviour may reflect the fact that the M–B was a successful polarity reversal. It has been suggested⁷ that the solid inner core largely controls the stability and hence the polarity of the dynamo field. Because the characteristic diffusion time for flux to escape the inner core is $\sim 10^3$ yr, it was suggested⁶ that the probability of a successful

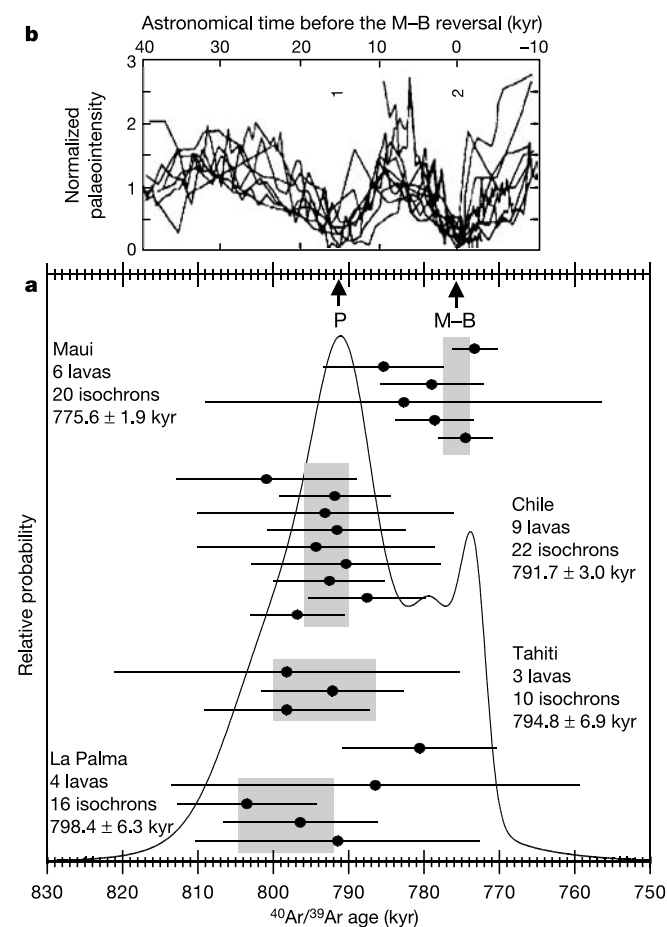


Figure 3 $^{40}\text{Ar}/^{39}\text{Ar}$ ages of 23 transitionally magnetized lava flows and palaeointensity records from 12 marine cores²⁵. **a**, Grey bands show the weighted mean age and 2σ uncertainty for each lava sequence. An exception is the uppermost lava from La Palma, which is younger than the underlying four lavas but coeval with lavas from Maui. The probability density curve indicates that these lavas span a minimum of 18 kyr. **b**, Normalized palaeointensity of sediments plotted so that the astronomical age of the M–B reversal matches the $^{40}\text{Ar}/^{39}\text{Ar}$ age of the Maui lavas (arrow at M–B). The precursory dip in intensity (arrow at P) is coeval with transitional lavas at Tahiti, La Palma and Chile.

reversal increases sharply as the duration of unstable outer-core field behaviour exceeds this value. The $^{40}\text{Ar}/^{39}\text{Ar}$ ages, indicating that dynamo instability lasted an order of magnitude longer than this diffusion time, offer the first radioisotopic observation supportive of this claim. It remains to be explored whether, to be successful, a reversal attempt requires an extended, multi-stage dynamo process like that seen not only for the M–B but also in other Cenozoic reversal records^{14,26}.

Examination of the modern-day non-axial dipole (NAD) field reveals that the global pattern of VGPs that would arise if the axial dipole field were to vanish is not random²⁷. In particular, a significant fraction of the Earth's surface encompassing the Pacific, including Tahiti, would experience field directions associated with south VGPs in and around Australia²⁷. Because the dynamo is blind to the sign of magnetic flux¹, the three most detailed records obtained from Tahitian lavas are wholly compatible with this simple model. Specifically, north VGPs associated with the R–R Punaruu^{15,20} and N–N Big Lost²⁷ events, as well as the N–R Jaramillo–Matuyama^{15,20} reversal, display a distinct preference for the Australasian region early in each record and contain a VGP cluster nearly identical to that of Tahitian M–B lavas shown in Fig. 3. Given the precise ages of the four M–B lava sequences, we can add another Tahitian result to this list. Along with the modern-day NAD-field analysis, these correspondences indicate that a mantle-controlled pattern of flux at the core surface might dominate the field configuration upon initiation of a dynamo instability when almost complete destruction of the axial dipole occurs. That transitional VGPs at this time on La Palma are found mainly elsewhere—over southern South America—lends support to a globally observed, largely non-dipolar early M–B field. Indeed, transitional M–B VGPs obtained from North Atlantic marine sediments^{9,28} indicate long-lived residences in this same locality.

Subsequent field behaviour associated with the actual polarity change might be more complex. Apparently, for the M–B transition, a significant fraction of the total 18 kyr of recorded instability elapsed before flux diffusion from the solid inner core allowed the reversal to proceed. □

Received 9 September 2004; accepted 31 January 2005; doi:10.1038/nature03431.

- Merrill, R. T. & McFadden, P. T. Geomagnetic polarity transitions. *Rev. Geophys.* **37**, 201–226 (1999).
- Laj, C., Mazaud, A., Weeks, R., Fuller, M. & Herrero-Bervera, E. Geomagnetic reversal paths. *Nature* **351**, 447 (1991).
- Glatzmaier, G. A., Coe, R. S., Hongre, L. & Roberts, P. H. The role of the Earth's mantle in controlling the frequency of geomagnetic reversals. *Nature* **401**, 885–890 (1999).
- Hoffman, K. A. Dipolar reversal states of the geomagnetic field and core-mantle dynamics. *Nature* **359**, 789–794 (1992).
- Coe, R. S. & Glen, J. M. G. The complexity of reversals. In *Timescales of the Paleomagnetic Field* (eds Channell, J. E. T., Kent, D. V., Lowrie, W. & Meert, J.) 221–232 (Am. Geophys. Un. Geophys. Monogr. 145, 2004).
- Gubbins, D. The distinction between geomagnetic excursions and reversals. *Geophys. J. Int.* **137**, F1–F3 (1999).
- Hollerbach, R. & Jones, C. A. Influence of the Earth's inner core on geomagnetic fluctuations and reversals. *Nature* **365**, 541–543 (1993).
- Laj, C., Guitton, S. & Kissel, C. Rapid changes and near-stationarity of the geomagnetic field during a polarity reversal. *Nature* **330**, 145–148 (1987).
- Channell, J. E. T. & Lehman, B. The last two geomagnetic polarity reversals recorded in high-deposition-rate sediment drifts. *Nature* **389**, 712–715 (1997).
- Kent, D. V. & Schneider, D. A. Correlation of paleointensity variation records in the Brunhes–Matuyama polarity transition interval. *Earth Planet. Sci. Lett.* **129**, 135–144 (1995).
- Channell, J. E. T. & Kleiven, H. F. Geomagnetic palaeointensities and astrochronologic ages for the Matuyama–Brunhes boundary and the boundaries of the Jaramillo Subchron: palaeomagnetic and oxygen isotope records from ODP site 983. *Phil. Trans. R. Soc. Lond. A* **358**, 1027–1047 (2000).
- Shackleton, N. J., Berger, A. & Peltier, W. R. An alternative astronomical calibration of the lower Pleistocene timescale based on ODP site 677. *Trans. R. Soc. Edinb. Earth Sci.* **81**, 251–261 (1990).
- Clement, B. Dependence of the duration of geomagnetic polarity reversal on site latitude. *Nature* **428**, 637–640 (2004).
- Hoffman, K. A. Transitional field behavior from southern hemisphere lavas: evidence for two-stage reversals of the geodynamo. *Nature* **320**, 228–232 (1986).
- Singer, B. S., Hoffman, K. A., Chauvin, A., Coe, R. S. & Pringle, M. S. Dating transitionally magnetized lavas of the late Matuyama Chron: Toward a new $^{40}\text{Ar}/^{39}\text{Ar}$ timescale of reversals and events. *J. Geophys. Res.* **104**, 679–693 (1999).
- Singer, B. S. *et al.* Ar/Ar ages of transitionally magnetized lavas on La Palma, Canary Islands, and the

- Geomagnetic Instability Timescale. *J. Geophys. Res. Solid Earth* **107**(B11), doi:10.1029/2001JB001613 (2002).
- Coe, R. S., Singer, B. S., Pringle, M. S. & Zhao, X. Matuyama–Brunhes reversal and Kamikatsura event on Maui: paleomagnetic directions, $^{40}\text{Ar}/^{39}\text{Ar}$ ages and implications. *Earth Planet. Sci. Lett.* **222**, 667–684 (2004).
- Brown, L., Singer, B. S., Pickens, J. & Jicha, B. Paleomagnetic directions and $^{40}\text{Ar}/^{39}\text{Ar}$ ages from the Tatara–San Pedro volcanic complex, Chilean Andes: Lava record of a Matuyama–Brunhes precursor? *J. Geophys. Res. Solid Earth* **109**(B12), doi:10.1029/2004JB003007 (2004).
- Love, J. J. & Mazaud, A. A database for the Matuyama–Brunhes magnetic reversal. *Phys. Earth Planet. Inter.* **103**, 207–245 (1997).
- Chauvin, A., Roperch, P. & Duncan, R. A. Records of geomagnetic reversals from volcanic islands of French Polynesia 2. Paleomagnetic study of a flow sequence (1.2 to 0.6 Ma) from the Island of Tahiti and discussion of reversal models. *J. Geophys. Res.* **95**, 2727–2752 (1990).
- Singer, B. S. & Pringle, M. S. Age and duration of the Matuyama–Brunhes geomagnetic polarity transition from $^{40}\text{Ar}/^{39}\text{Ar}$ incremental heating analyses of lavas. *Earth Planet. Sci. Lett.* **139**, 47–61 (1996).
- Brown, L., Pickens, J. & Singer, B. Matuyama–Brunhes transition recorded in lava flows of the Chilean Andes: Evidence for dipolar fields during reversals. *Geology* **22**, 299–302 (1994).
- Renne, P. R. *et al.* Inter-calibration of standards, absolute ages and uncertainties in $^{40}\text{Ar}/^{39}\text{Ar}$ dating. *Chem. Geol.* **145**, 117–152 (1998).
- Tauxe, L., Herbert, T., Shackleton, N. J. & Kok, Y. S. Astronomical calibration of the Matuyama–Brunhes boundary: Consequences for magnetic remanence acquisition in marine carbonates and the Asian loess sequences. *Earth Planet. Sci. Lett.* **140**, 133–146 (1996).
- Hartl, P. & Tauxe, L. A precursor to the Matuyama/Brunhes transition-field instability as recorded in pelagic sediments. *Earth Planet. Sci. Lett.* **138**, 121–135 (1996).
- Prévot, M., Mankinen, E. A., Gromme, C. S. & Coe, R. S. How the geomagnetic field vector reverses polarity. *Nature* **316**, 230–234 (1985).
- Hoffman, K. A. & Singer, B. S. Regionally recurrent paleomagnetic transitional fields and mantle processes. In *Timescales of the Paleomagnetic Field* (eds Channell, J. E. T., Kent, D. V., Lowrie, W. & Meert, J.) 223–243 (Am. Geophys. Un. Geophys. Monogr. 145, 2004).
- Clement, B. M. & Kent, D. V. Geomagnetic polarity transition records from five hydraulic piston cores sites in the North Atlantic. In *Init. Rep. Deep Sea Drilling Project* (eds Ruddiman, W. F. *et al.*) **94**, 831–852 (US Government Printing Office, Washington, 1987).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Pickens, M. Relle, A. Battle, L. Powell and R. Allen for assistance with field work, argon and palaeomagnetic analyses, and graphics. This study was supported by the US NSF.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.S. (bsinger@geology.wisc.edu).

Sex increases the efficacy of natural selection in experimental yeast populations

Matthew R. Goddard*, H. Charles J. Godfrey & Austin Burt

NERC Centre for Population Biology and Department of Biological Sciences, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK

* Present address: School of Biological Sciences, The University of Auckland, Private Bag 92019, Auckland, New Zealand

Why sex evolved and persists is a problem for evolutionary biology, because sex disrupts favourable gene combinations and requires an expenditure of time and energy¹. Further, in organisms with unequal-sized gametes, the female transmits her genes at only half the rate of an asexual equivalent (the twofold cost of sex)². Many modern theories that provide an explanation for the advantage of sex incorporate an idea originally proposed by Weismann more than 100 years ago: sex allows natural selection to proceed more effectively because it increases genetic variation^{3–5}. Here we test this hypothesis, which still lacks robust empirical support, with the use of experiments on yeast populations. Capitalizing on recent advances in the molecular biology of recombination in yeast, we produced by genetic manipulation

strains that differed only in their capacity for sexual reproduction. We show that, as predicted by the theory, sex increases the rate of adaptation to a new harsh environment but has no measurable effect on fitness in a new benign environment where there is little selection.

The Weismann effect is expected whenever fitness correlations between loci are negative. Under these circumstances, the net effect of sex is to bring together favourable mutations, separate favourable mutations from harmful mutations, and concentrate harmful mutations, all of which can increase the efficacy of selection. Negative fitness correlations can arise from epistatic interactions between loci and through the effects of finite population size⁴. In diploid organisms, sex can also increase the variance of fitness by the segregation of alleles at individual loci or because of selection on the transient haploid phase^{6,7}.

If the Weismann effect is important, then adaptation to new evolutionary challenges should proceed faster in sexual populations than in asexual populations. Experiments with yeast and *Chlamydomonas*^{8–11}, along with work on other taxa^{12,13}, have largely supported the hypothesis, although studies so far are subject to several possible criticisms. Sexual reproduction in microbial model systems is normally manipulated by starvation, which means that asexual and sexual replicates experience different selection regimes. Starvation in both *Chlamydomonas* and yeast is known to increase mutation rates^{14,15}. One study with *Chlamydomonas* avoided such treatment differences, but at the expense of the association of sex with large-scale genetic introgression¹⁰. These problems will tend to increase genetic variation, so these manipulations might have significant effects on the course of evolution that are independent of sexual reproduction. Advances in our understanding of the molecular mechanisms of meiosis in yeast allow us to circumvent these problems by genetically engineering an asexual strain that can be treated identically to the sexual wild type. This allows us to study the net effect of sex itself on the rate of adaptation.

Yeast cells grow vegetatively when supplied with sufficient nutrients; however, if starved, diploids undergo sporulation (meiosis) and produce four haploid spores encapsulated within an ascus¹⁶. Diploid yeast are heterozygous (a/α) at the mating-type locus (*MAT*); the ascus therefore contains two spores of each mating type. When conditions improve, these spores germinate and mate with a spore of the opposite mating type. Yeast are isogamous (each spore contributes equally to the resulting diploid zygote) and so this system has no twofold cost of sex. Apomictic strains of yeast that produce pairs of diploid spores (dyads) instead of four haploid spores (tetrads) have occasionally been recovered from natural populations¹⁷.

We constructed an apomictic (asexual) strain of yeast that differs from the wild type in its capacity for recombination, random assortment and syngamy by deleting two genes required for normal recombination and meiosis. *SPO11* encodes an endonuclease that initiates crossover events by making double-stranded breaks in chromosomes: in its absence meiotic recombination does not occur¹⁸. *SPO13* determines whether a cell goes through one or two meiotic divisions by altering the sister-chromatid cohesion process¹⁹; in its absence only the second non-reductive meiotic division is achieved, resulting in the production of two diploid spores²⁰. Because chiasmata are required for the stabilization of chromosome segregation, non-functional mutations of *SPO11* would normally lead to aberrant chromosomal segregation, but this phenotype is rescued if *SPO13* is also non-functional; this leaves the asexual double mutant fully fertile, producing diploid spores that are genetically identical to the parent cell (refs 18, 21 and see below). We deleted both of these genes by using standard yeast engineering techniques, and replaced *SPO13* with the bacterial *kanMX4* G418 resistance gene to provide a marker to distinguish the asexual strain²². The engineered asexual strain is thus not only

able to sporulate but also genetically identical to the sexual strain except at these two loci.

Weismann's hypothesis predicts that sex will lead to an increase in mean fitness when there is selection, but not otherwise. We tested this hypothesis by comparing the rate of adaptation of replicate sexual and asexual yeast populations (eight each) when exposed to one of two environments: a new benign chemostat environment in which growth was limited by glucose concentration (0.08%), or a harsher chemostat environment with the same glucose concentration but in which the temperature was raised from 30 to 37 °C, and where osmolarity was elevated from 0 to 0.2 M NaCl. The carrying capacities were about 3×10^7 and about 3×10^6 cells ml⁻¹ under the benign and harsh conditions respectively, which reflected the relative stresses experienced by these yeast in the two environments.

SPO13 and *SPO11* are expressed only during meiosis^{18,20}, and the *kanMX4* marker has previously been shown to have a negligible fitness effect in yeast when propagated in glucose-limited chemostats²³. However, to rule out any unexpected fitness effects of engineering we allowed the ancestral sexual and asexual strains to compete under both the benign and harsh environments. The average intrinsic exponential growth rate (fitness) advantage of the sexual strain over the asexual strain in both environments was similar and very near zero (benign, 0.003 ± 0.0016 (mean $\pm$ s.e.m.); harsh, 0.003 ± 0.0023 ; $n = 8$). We also checked that the manipulation had not affected sporulation efficiency and found that the proportion of cells that produced spores was very similar in both the ancestral sexual and engineered asexual strains (sexual, $76 \pm 4\%$ (mean $\pm$ s.e.m.); asexual, $73 \pm 4\%$; t -test, $P = 0.62$).

Replicate experimental populations of the sexual and asexual strains were each initiated from a single colony and subjected to regular cycles of vegetative growth punctuated by sporulation. Each vegetative growth phase lasted 4 days and took place in about 33 ml of medium in a shaking chemostat. Chemostat generation times (determined by nutrient in-flow rate) were set at about 2.2 h and about 4.0 h for the ancestral strains in the benign and harsh environments, respectively. This meant that in the benign environment there were about 50 mitotic generations between each sporulation episode, and in the harsh environment about 25. To induce sporulation the medium was replaced with a 1% potassium acetate solution; all populations were then incubated without the addition or removal of medium for 7 days at 30 °C (ref. 24), and subsequently removed from the chemostat vessels. Mating after germination normally occurs within asci, and yeast would therefore naturally inbreed; however, to promote outbreeding and to destroy unsporulated cells, the yeast cultures had their asci and cell walls softened with the enzymes β -glucuronidase and lyticase. They were then placed in a 1% solution of the surfactant sodium dodecyl sulphate to lyse unsporulated cells, followed by sonication to release and thoroughly mix the spores. Germination was triggered when the yeast were returned to their normal growth medium, after which mating occurred in the sexual strains. We estimated the level of outbreeding obtained under these conditions by crossing the sexual strain, a uracil auxotroph, with a strain that was genetically identical except that it was a lysine auxotroph. We found that $86 \pm 3\%$ (mean $\pm$ s.e.m., $n = 8$) of the progeny were outcrossed. Both experiments were run for about 300 vegetative generations, equivalent to five (benign environment) or ten (harsh environment) cycles of growth and sporulation.

We estimated the changes in fitness of the populations over the course of the experiment by allowing samples from the replicate populations to compete against an ancestral strain (yeast can be stored indefinitely at -80 °C in glycerol) under conditions identical to those in the experiment. In these competition assays we determined the proportion of each strain using selective plating at the start and then again after an average of 30 mitotic generations. A

measure of relative fitness under continuous growth was obtained by subtracting the initial from the final natural logarithms of the ratios, and the relative fitness per mitotic generation was found by dividing this figure by the number of generations in each particular competition assay. Confidence limits were calculated taking into account the binomial sampling variance.

We were unable to detect any change in the fitness of either the sexual population or the asexual population in the benign environment (Fig. 1). The overall average cell density neither differed significantly between strains nor changed significantly over the course of the experiment.

In contrast with the experiment in the benign environment, the relative fitness of yeast populations exposed to the harsh environment increased markedly over the course of the experiment. We define relative fitness here as the natural logarithm of population growth rate relative to the ancestral strain (which therefore has a relative fitness of zero). In the asexual lines, relative fitness increased from 0 to 0.59, whereas for the sexual lines the increase was from 0 to 0.68. These figures are equivalent to increases in population growth rate of 80% and 94%, respectively. Inspection of the data and exploratory analysis indicated that fitness was beginning to reach a plateau towards the end of experiment and was therefore analysed with a nonlinear mixed-effect model²⁵,

$$w_i(t) = a_{1i}[1 - \exp(-A_{2i}t)] + \varepsilon_i(t)$$

where $w_i(t)$ is the relative fitness of replicate i at time t . The nonlinear function, which is constrained to be zero at time $t = 0$, is specified by replicate-specific asymptote (a_{1i}) and rate ($a_{2i} = \ln(A_{2i})$) parameters, each of which in the full model is described by the sum of statistical random and fixed components. The random component reflects differences between replicates, and the fixed component reflects the treatment effects of sexual versus asexual lines. $\varepsilon_i(t)$ is the within-group variance assumed to be independently and normally distributed with zero mean. After simplification of the model, the random effect was confined to the asymptote parameter a_1 and the fixed effect was restricted to the rate parameter a_2 (although an alternative model with the

treatment effect on a_1 was nearly as informative). We tested for the significance of sexual versus asexual reproduction by removing the treatment effect, which was highly significant (log-likelihood ratio 12.6, $P = 0.0004$). The data and fitted model are shown in Fig. 1.

In the new benign environment there was no detectable adaptation over the course of the experiment, and thus no detectable selection for beneficial mutations. There would still have been purifying selection clearing deleterious mutations as they arose. In other experiments it has been estimated that deleterious mutations decrease mean fitness in yeast by about 0.002% in each generation²⁶. At mutation–selection balance, selection need only increase fitness by this amount to prevent the accumulation of maladaptive variants. This is a very small value, and any difference there might have been in the efficacy with which sexual and asexual populations cleared their mutations would have been undetectable in our experiment.

By contrast, in the harsh environment, relative fitness increased markedly for both the asexual and sexual populations; after 100 generations the intrinsic rate of increase in the asexual populations exceeded the ancestor by 0.3, but the equivalent figure for the sexual populations was 0.4. Applying Fisher's Fundamental Theorem of Natural Selection this indicates that in the first 100 generations of the experiment there must have been genetic variance in fitness, upon which selection could act, of about 0.003 and 0.004 per generation in the asexual and sexual populations, respectively. This is consistent with the expectation from Weismann's hypothesis that the maintenance of sex is associated with increased variance in fitness. The difference in fitness between the sexual and asexual populations after 100 generations, and indeed throughout most of the experiment, is about 0.1. Although this may not seem a very great advantage, the geometric growth process underlying it quickly leads to large differences in cell numbers. Thus in the 25 mitotic generations between episodes of sex in our experiments, a sexual individual could expect to leave about 12-fold ($e^{0.1 \times 25}$) as many descendants as an asexual individual. We can also ask whether the rates of evolution in our experiments are unrealistically high. This does not seem to be so, because the 0.3–0.4% variation in fitness we observed is in the lower range of estimates for natural populations of various plants and animals (0.1–30%, with typical values likely to be between 1% and 10%)²⁷.

Recombination increases the efficacy of natural selection when fitness correlations between loci are negative^{4,5}. There are at least two non-exclusive reasons why negative correlations might have occurred in our experiments: epistatic interactions between new mutations, and the effects of finite population size. Finite population size is significant in circumstances in which sex allows adaptive mutations that have occurred in different lineages to be brought together and to increase in frequency simultaneously, rather than their being forced to spread sequentially as they would in an asexual population²⁸. Further genetic work to isolate and characterize the mutations responsible for the adaptation to the harsh environment is needed to distinguish between these possibilities.

Meiotic recombination has been seen by many theoreticians as the most important aspect of sex², but other effects of sex might also have contributed to the increased rate of adaptation. In diploids the segregation of alleles at a locus can be associated with increased variance in fitness, particularly in populations with some inbreeding (as occurred in our experiments)⁷. In addition, the asexual strains would not have had a brief haploid phase upon which selection could act⁶. In some models of selection with fluctuating epistasis it is possible for sex to increase mean fitness directly, without an effect on the variance²⁹. We think that this is unlikely to explain our results: although the populations were exposed to two separate environments (one for vegetative growth and one for sporulation), sex occurred every cycle rather than between each

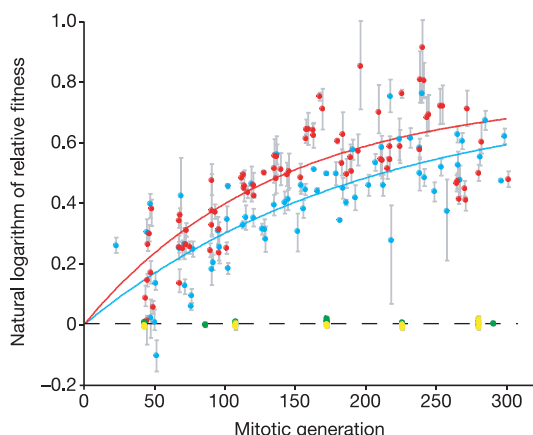


Figure 1 The change in natural logarithm of fitness of asexual and sexual populations of yeast in benign and harsh environments. Points show fitness measurements for individual populations with twice log-likelihood error bars (these approximate 95% confidence limits); the error bars for the benign treatment are plotted but are mostly too small to be discriminated. The fitted model for the harsh environment is plotted for asexual (blue) and sexual (red) treatments (parameters: $a_1 = 0.761$, $a_2(\text{asexual}) = -5.287$, $a_2(\text{sexual}) = -4.901$). Yellow symbols, asexual strains in the benign environment; green, sexual in the benign environment; blue, asexual in the harsh environment; red, sexual in the harsh environment.

transition as the model requires. Finally, it is conceivable that the engineered knockouts in the asexuals had some pleiotropic effect of reducing mutation rates.

Our results indicate that sexual reproduction can provide a selective advantage during adaptation to new environments, and these data are consistent with Weismann's ideas. They are an advance on earlier studies because by manipulating sexual status genetically we have mimicked the natural situation more closely and have excluded a variety of possible confounding factors. In extrapolating from experiments with yeast and other single-celled organisms to multicellular organisms, at least two points need to be considered. First, yeast gametes are of equal size (isogamous) and hence the classical twofold cost of sex does not apply to these organisms. There are still costs to sex (see above), but they will be less than in anisogamous species. Second, empirical studies of deleterious mutations in yeast indicate that their effects on fitness are significantly lower than in 'higher' eukaryotes. Therefore one of the major postulated advantages of sex, the removal of epistatically interacting deleterious mutations, will be less important. Population experiments with fast-replicating microorganisms have been very valuable in testing different ideas about the maintenance of sex, and a challenge now is to understand the nature of the mutations that underlie adaptation and to extend these techniques to larger plants and animals. □

Methods

Additional details are given in Supplementary Methods.

Strain construction

We subcultured a Y55-derived haploid α , *ho*, *ura3Δ* colony and switched mating types²⁴ to yield isogenic MAT a and α lines. We employed polymerase-chain-reaction-based gene targeting to replace *SPO13* with the *kanMX4* module²² in the α strain and to delete the *SPO11* locus from the a strain (by using two-step uracil plus/minus selection). Haploids were mated to produce a strain heterozygous at only the *SPO11*, *SPO13* and *MAT* loci; this strain sporulated normally and produced four viable spores. MGA1 {a, *ho*, *ura3Δ*, *spo11Δ*, *spo13Δ::kanMX4*} and MGS1 { α , *ho*, *ura3Δ*} were isolated from one tetrad, and MGA2 { α , *ho*, *ura3Δ*, *spo11Δ*, *spo13Δ::kanMX4*} and MGS2 {a, *ho*, *ura3Δ*} from another. Mating yielded strains with identical histories of genetic engineering and isogenic at all but the *SPO11* and *SPO13* loci. Lysine auxotrophs were selected on 5% α -aminoadipic acid²⁴.

Media and growth

The basic mitotic growth medium consisted of 0.17% yeast nitrogen base without C, N or amino acids (Sigma), 0.5% ammonium sulphate, 0.08% glucose, 20 mg l⁻¹ uracil (Sigma). Continuous culture occurred in a replicated chemostat unit (Infors-HT); flasks were shaken at 125 r.p.m. with independent aeration. The dilution rate averaged 0.31 h⁻¹ for the benign treatment and 0.17 h⁻¹ for the harsh treatment. Sporulation medium comprised 1% potassium acetate and 5 mg l⁻¹ uracil. 'Outcrossing solution' (10 ml; 40 U ml⁻¹ lyticase (Sigma), 1,107 U ml⁻¹ β -glucuronidase (Sigma), 50 mM dithiothreitol (Sigma); all filtered at 0.2 μ m) was added to the cultures and incubated overnight at 37 °C, with shaking. Cross-contamination was checked by using the G418 resistance marker and by visual inspection of the spore structures; it was found only in one case when a sexual population invaded an asexual population in the harsh treatment. Samples (25 ml) were taken from the outflow just before sporulation; they were condensed and stored at -80 °C in 15% v/v glycerol for subsequent fitness estimation.

Outcrossing assay

Eight populations were initiated with approximately equal frequencies of the diploid lys⁻ and ura⁻ strains. Control populations with single or haploid strains were also initiated, and all were grown in chemostats for about ten generations, sporulated, and subjected to the outcrossing treatment described above. The amino-acid requirements of the progeny allowed the proportion of each genotype to be estimated: the mean inbreeding coefficient (Wright's *F*) of replicate populations (more than 1,500 sampled colonies) was 0.13 $\pm$ 0.03 (mean $\pm$ s.e.m.). No haploid (unsporulated) cells survived the treatment.

Fitness assay

Difference in resistance to G418 toxin allowed us to distinguish between competitors. Confidence limits were placed around each fitness estimate with a binomially based maximum-likelihood model. Replicates of individual competition assays converged on very similar estimates (within support limits of one another). Initial competitions were against the ancestor; by generation 150 the fitness of the derived populations under the harsh treatment had increased to the point at which ancestral strains were driven below detectable frequencies in competition assays. We therefore marked a derived sexual strain with G418 resistance and, after estimating the fitness effects of marking (-0.082 $\pm$ 0.0090, mean $\pm$ s.e.m., *n* = 8), used this as a competitive standard. The

competitions were conducted over an average of 12 (harsh) and 50 (benign) generations. In every case, final fitness estimates were adjusted to take into account the relative fitness of the standard competitor; in this way all fitness estimates were rendered directly comparable.

Statistical analyses

Data were analysed with the R statistical package (<http://www.r-project.org/>) on the basis of the model shown in the main text. If w_{ij} is the *j*th measurement (at time t_{ij}) of the fitness of line *i*, then the statistical model can be written

$$w_{ij} = a_{1i}(1 - \exp(-\exp(a_{2i}t_{ij}))) + \varepsilon_{ij},$$

$$\begin{bmatrix} a_{1i} \\ a_{2i} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} = \beta_k + b_i,$$

$$b_i \sim N(0, \Psi), \quad \varepsilon_{ij} \sim N(0, \sigma^2).$$

The fixed effects, β_k , represent the average value of a sexual ($k = 1$) or asexual ($k = 2$) population. The random effects, b_i , are normally distributed with a mean of 0 and a variance-covariance matrix Ψ , and are assumed to be constant in time and independent for different lines. The within-group variance, ε_{ij} , is normally distributed with mean zero and variance σ^2 , and is independent for different *i*, *j* and independent of the random effects. Following ref. 25 we fitted the full random effects model and examined the fitted variance-covariance matrix to see whether all terms were necessary. The random effect b_{2i} was small and correlated with b_{1i} ; its removal made little difference to the fit of the model. We therefore restricted the random effect to a_1 . The data could be explained almost equally well by assuming that the treatment influenced a_1 or a_2 , with an effect on the latter providing a slightly better fit. Allowing both parameters to depend on treatment did not significantly improve the fit (or produce a better Akaike's Information Criteria (AIC) score) compared with either single-parameter model.

Received 22 November 2004; accepted 27 January 2005; doi:10.1038/nature03405.

1. Bell, G. *The Masterpiece of Nature* (Univ. California Press, Berkeley, 1982).
2. Michod, R. E. & Levin, B. R. *The Evolution of Sex: An Examination of Current Ideas* (Sinauer, Sunderland, Massachusetts, 1988).
3. Weismann, A. *The Evolution Theory* (Edward Arnold, London, 1904).
4. Barton, N. H. & Charlesworth, B. Why sex and recombination? *Science* **281**, 1986–1990 (1998).
5. Burt, A. Sex, recombination, and the efficacy of selection—was Weismann right? *Evolution* **54**, 337–351 (2000).
6. Bruggeman, J., Debets, A. J. M., Wijngaarden, P. J., deVisser, J. A. G. M. & Hoekstra, R. F. Sex slows down the accumulation of deleterious mutations in the homothallic fungus *Aspergillus nidulans*. *Genetics* **164**, 479–485 (2003).
7. Otto, S. P. & Lenormand, T. Resolving the paradox of sex and recombination. *Nature Rev. Genet.* **3**, 252–261 (2002).
8. Birdsall, J. & Wills, C. Significant competitive advantage conferred by meiosis and syngamy in the yeast *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **93**, 908–912 (1996).
9. Greig, D., Borts, R. H. & Louis, E. J. The effect of sex on adaptation to high temperature in heterozygous and homozygous yeast. *Proc. R. Soc. Lond. B* **265**, 1017–1023 (1998).
10. Colegrave, N. Sex releases the speed limit on evolution. *Nature* **420**, 664–666 (2002).
11. Kaltz, O. & Bell, G. The ecology and genetics of fitness in *Chlamydomonas*. XII. Repeated sexual episodes increase the rates of adaptation to novel environments. *Evolution* **56**, 1743–1753 (2002).
12. Rice, W. R. Experimental effects of the adaptive significance of sexual recombination. *Nature Rev. Genet.* **3**, 241–251 (2002).
13. Rice, W. R. & Chippindale, A. K. Sexual recombination and the power of natural selection. *Science* **294**, 555–559 (2001).
14. Goho, S. & Bell, G. Mild environmental stress elicits mutations affecting fitness in *Chlamydomonas*. *Proc. R. Soc. Lond. B* **267**, 123–129 (2000).
15. Marini, A., Matmati, N. & Morpurgo, G. Starvation in yeast increases non-adaptive mutation. *Curr. Genet.* **35**, 77–81 (1999).
16. Berry, D. R. *The Biology of Yeasts* (Edward Arnold, London, 1982).
17. Grewal, N. S. & Miller, J. J. Formation of asci with two diploid spores by diploid cells of *Saccharomyces*. *Can. J. Microbiol.* **18**, 1897–1905 (1972).
18. Klapholz, S., Waddell, C. S. & Esposito, R. E. The role of the *SPO11* gene in meiotic recombination in yeast. *Genetics* **110**, 187–216 (1985).
19. Shonn, M. A., McCarroll, R. & Murray, A. W. Spo13 protects meiotic cohesin at centromeres in meiosis I. *Genes Dev.* **16**, 1659–1671 (2002).
20. Wang, H., Frackman, S., Kowalisyn, J., Esposito, R. E. & Elder, R. Developmental regulation of *SPO13*, a gene required for separation of homologous chromosomes at meiosis I. *Mol. Cell. Biol.* **7**, 1425–1435 (1987).
21. Steele, D. F., Morris, M. E. & Jinks-Robertson, S. Allelic and ectopic interactions in recombination-defective yeast strains. *Genetics* **127**, 53–60 (1991).
22. Wach, A. et al. PCR-based gene targeting in *S. cerevisiae*. *Methods Microbiol.* **26**, 67–81 (1998).
23. Baganz, F., Hayes, A., Marren, D., Gardner, D. C. J. & Oliver, S. G. Suitability of replacement markers for functional analysis studies in *Saccharomyces cerevisiae*. *Yeast* **13**, 1536–1573 (1997).
24. Burke, D., Dawson, D. & Stearns, T. *Methods in Yeast Genetics. A Cold Spring Harbor Laboratory Course Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2000).
25. Pinheiro, J. C. & Bates, D. M. *Mixed-effects Models in S and S-PLUS* (Springer, New York, 2000).
26. Zyl, C. & DeVisser, J. A. G. M. Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. *Genetics* **157**, 53–61 (2001).
27. Burt, A. The evolution of fitness. *Evolution* **49**, 1–8 (1995).
28. Muller, H. J. Some genetic aspects of sex. *Am. Nat.* **66**, 118–138 (1932).
29. Barton, N. A general model for the evolution of recombination. *Genet. Res.* **65**, 123–144 (1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Rees for comments on statistical analyses. This work was supported by the UK Natural Environment Research Council (NERC).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.R.G. (m.goddard@auckland.ac.nz).

Functional consequences of a *CKIδ* mutation causing familial advanced sleep phase syndrome

Ying Xu^{1*}, Quasar S. Padiath^{1*}, Robert E. Shapiro², Christopher R. Jones³, Susan C. Wu¹, Noriko Saigoh¹, Kazumasa Saigoh^{1†}, Louis J. Ptáček^{1,4} & Ying-Hui Fu¹

¹Department of Neurology, University of California, San Francisco, San Francisco, California 94143-2922, USA

²Department of Neurology, College of Medicine, University of Vermont, Burlington, Vermont 05405, USA

³Department of Neurology, University of Utah, Salt Lake City, Utah 84132-2305, USA

⁴Howard Hughes Medical Institute, UCSF, San Francisco, California 94143-2922, USA

* These authors contributed equally to this work

† Present address: Department of Neurology, Kinki University School of Medicine, Osaka 589-8511, Japan

Familial advanced sleep phase syndrome (FASPS) is a human behavioural phenotype characterized by early sleep times and early-morning awakening¹. It was the first human, mendelian circadian rhythm variant to be well-characterized, and was shown to result from a mutation in a phosphorylation site within the casein kinase I (CKI)-binding domain of the human *PER2* gene. To gain a deeper understanding of the mechanisms of circadian rhythm regulation in humans, we set out to identify mutations in human subjects leading to FASPS. We report here the identification of a missense mutation (T44A) in the human *CKIδ* gene, which results in FASPS. This mutant kinase has decreased enzymatic activity *in vitro*. Transgenic *Drosophila* carrying the human *CKIδ-T44A* gene showed a phenotype with lengthened circadian period. In contrast, transgenic mice carrying the same mutation have a shorter circadian period, a phenotype mimicking human FASPS. These results show that *CKIδ* is a central component in the mammalian clock, and suggest that mammalian and fly clocks might have different regulatory mechanisms despite the highly conserved nature of their individual components.

Phosphorylation has a central role in the regulation of circadian clocks^{2–5}. Multiple points of regulation have been proposed, including the nuclear import and export of circadian proteins, and the active/inactive states of these proteins, but the details of this system are not well understood^{6–8}. In *Drosophila*, *Per* and *Tim* are two core proteins of the circadian clock. Doubletime (*Dbt*) and casein kinase II (*CKII*) phosphorylate *Per*, and *Shaggy* (*Sgg*) phosphorylates *Tim*^{9–13}. The Syrian hamster *tau* mutation was identified in the gene for casein kinase Iε (*CKIε*), a mammalian homologue of *Drosophila dbt*¹⁴. The known mutations in *dbt*, *CKII*, and *CKIε* lead to hypophosphorylation *in vitro*^{11,12,14,15}. However, the different mutations can give rise to different phenotypes *in vivo*, with longer or shorter circadian periods (τ).

FASPS is a behavioural phenotype manifest by early sleep times, early-morning awakening, and a short τ (ref. 1). It was shown to

result from mutation in a phosphorylation site within the CKI-binding domain of the human (h) *PER2* protein. To gain a deeper understanding of the mechanisms of human circadian regulation, we set out to identify mutations in human subjects that lead to FASPS. We report here the identification of a mutation in the human *CKIδ* gene (also known as *CSNK1δ*), which causes FASPS in humans. This mutation results in reduced kinase activity *in vitro* and leads to a shorter circadian period in mice, but a longer period in *Drosophila*.

The proband of this study was noted by one of the authors (R.E.S.) to have FASPS. Further evaluation of this subject's family revealed autosomal-dominant transmission of the behavioural trait. Fifteen family members, spanning three generations, were separately interviewed for their typical work and vacation sleep-wake schedules by two of the authors (C.R.J. and R.E.S.). Five individuals definitely affected with FASPS (age range 20–65 yr, mean 41 yr) were identified (Fig. 1a). In the absence of competing psycho-social demands, both the average sleep onset time (18:12 ± 1.4 h versus 23:24 ± 1.1 h) and final wake time (04:06 ± 0.7 h versus 08:00 ± 1.6 h) of these subjects were significantly earlier ($P < 0.0001$) than the nine unaffected family members. Affected individuals reported the onset of FASPS somewhere between early childhood to the mid-teen years.

Clinical features or a history of depression were found in three of the FASPS subjects, two of who had mildly elevated Beck Depression Inventory scores of 12 (ref. 16). A tendency towards winter depression was reported by a fourth FASPS subject. However, early-morning awakening due to depression was considered an unlikely explanation for FASPS in this family because the affected individuals showed little difficulty initiating or maintaining sleep, and had good energy levels in the morning (subjectively assessed). In addition, the young age of FASPS onset in these individuals was well below the age at which the typical early-morning awakening develops in depression¹⁷.

While screening candidate genes for circadian mutations (Supplementary Information), an A-to-G change was identified in the DNA sequence of *CKIδ*, which causes a threonine-to-alanine alteration at amino acid 44 in the protein (Fig. 1b). This

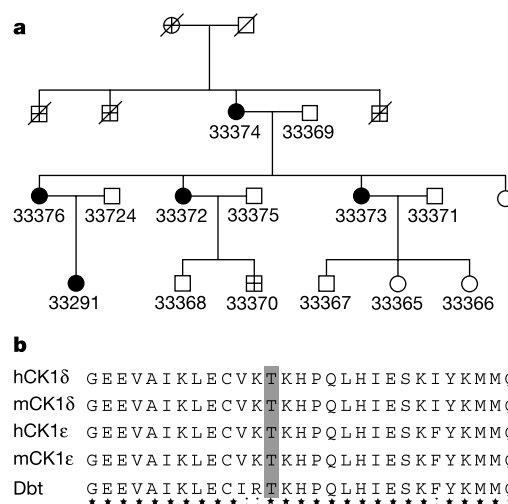


Figure 1 *CKIδ-T44A* FASPS pedigree and the amino acid alignment around the mutation. **a**, FASPS kindred 5231. Circles represent women, squares denote men, filled circles and squares show affected individuals; empty circles and squares show unaffected individuals. The individual marked with a cross is 'probably affected' but was conservatively classified as unknown. Diagonal lines across symbols indicate deceased individuals. **b**, Alignments for *Drosophila* *Dbt* and mouse (m) and human (h) *CKIδ* and *CKIε* proteins. The T44A mutation is highlighted.

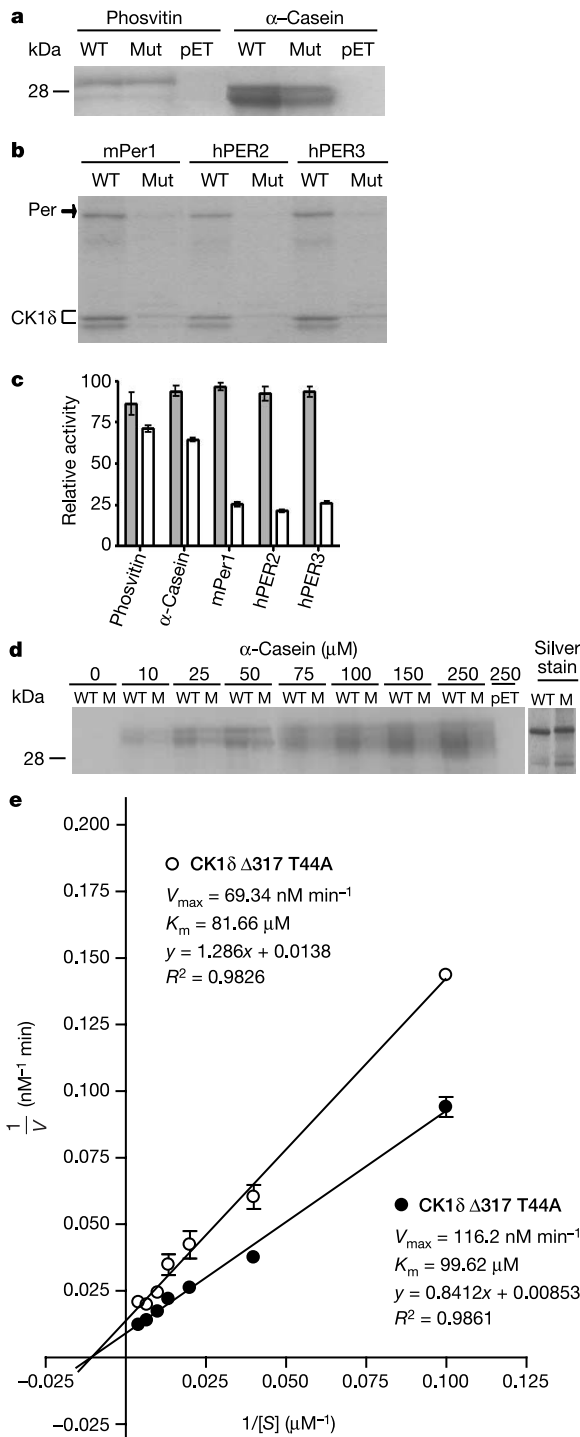


Figure 2 Biochemical characterization of CKIδ-T44A. **a, b**, *In vitro* kinase assays of wild-type and mutant CKIδΔ317 using the generic substrates phosvitin (100 μM) or α-casein (100 μM) (**a**), or *in vitro* translated circadian substrates, mPer1, hPER2 and hPER3 (**b**). pET, empty vector control. **c**, Quantitative analysis of kinase assays from **a** and **b**. Relative activity is defined as the percentage of activity relative to the highest measured activity for the same substrate. The experiment was performed in triplicate. Error bars indicate s.d. **d**, Kinetic analysis of wild-type (WT) and mutant (M) CKIδΔ317 using varying concentrations of α-casein. An empty vector control (pET) was included. One autoradiogram representative of nine experiments is shown. **e**, Lineweaver–Burk (double-reciprocal) plot of data derived from **d**. Each point represents the average ± s.e.m. from three experiments.

mutation (CKIδ-T44A) occurs at a residue that is conserved from mammalian CKIs through *Drosophila* CKI (also known as Dbt), and co-segregates with the FASPS phenotype in this family. This particular change was not found in over 250 control DNA samples.

We expressed the wild-type and mutant carboxy-terminal-truncated CKIδΔ317 (ref. 18) in a bacterial expression system (Supplementary Information). *In vitro* kinase assays with these bacterially expressed proteins and exogenous substrates (phosvitin and α-casein) showed that CKIδ-T44A has decreased activity compared with wild-type protein (Fig. 2a). Similar results were obtained with circadian-relevant substrates, PER1–3 (Fig. 2b). Interestingly, the defect in phosphorylation is greater when using PER polypeptides compared with α-casein or phosvitin (Fig. 2c). This observation suggests that CKIδ-T44A may have different effects on different substrates. Michaelis–Menten kinetic analyses were performed using wild-type and CKIδ-T44A enzyme (Fig. 2d). The Michaelis constant (K_m) and maximum reaction velocity (V_{max}) for both wild-type and CKIδ-T44A were calculated from nonlinear regressions of untransformed data. The resulting Lineweaver–Burk plot is shown in Fig. 2e. The V_{max} for CKIδ-T44A was 60% of the wild-type V_{max} , but the difference between K_m values is less dramatic (K_m for CKIδ-T44A is 82% of wild type).

We introduced the human CKIδ gene carrying the T44A mutation into flies in order to examine the effect on circadian activity (Supplementary Information). For accurate comparison, transgenic flies were generated containing either the wild-type or the mutant *hCKIδ* genes driven by the UAS–GAL4 system, using *tim*-UAS-*gal4* for expression in the lateral neurons, which are the central pacemaker cells of the *Drosophila* clock¹⁹. Three independent lines for each wild-type and mutant UAS-*hCKIδ* were crossed to *tim*-UAS-*gal4* flies and their locomotor activity was studied in constant darkness (DD) after entrainment for 3–4 days under 12 h light–12 h dark (LD) conditions (Supplementary Fig. 1a). Transgenic lines expressing the human CKIδ gene showed a longer circadian period length than the wild-type lines and undriven lines (which contain the transgene but are not crossed to *gal4* lines, and so do not express CKIδ). In all the *tim*-UAS-*gal4*/UAS-*hCKIδ* lines studied, transgenic flies expressing the mutant form of the *hCKIδ* gene had a significantly longer period than the flies with the wild-type *hCKIδ* gene (Table 1).

To ensure that the elongation of period seen in the *hCKIδ*-T44A transgenic flies was not due to increased expression of the transgene, we carried out semi-quantitative polymerase chain reaction with reverse transcription (RT–PCR) analysis using messenger RNA from fly heads (Supplementary Fig. 1b). The wild-type fly line (E25), which had a higher expression of the *hCKIδ* gene than mutant lines (H3 and H10), had a shorter period than all the mutant lines. This suggests that the mutation has functional consequences, as the mutant transgene leads to greater period elongation than the wild-type transgene. We then introduced the mutation at the conserved T44 residue of *dbt*, the *Drosophila* homologue of *hCKIδ*. Flies expressing the wild-type *dbt* gene driven by *tim*-UAS-*gal4* showed a period of 24.97 h. Flies expressing *dbt* with the *hCKIδ*-T44A mutation showed period elongation to 25.56 h (Supplementary Table 1). This period elongation is similar to observations when the human CKIδ-T44A mutant gene was expressed in flies. As an additional control, we also expressed the *dbt*^h allele, which is caused by a Thr–Ile mutation at the same residue as the FASPS *hCKIδ* mutation. These flies showed an extreme elongation of circadian period to 29.45 h, similar to the period seen in the *dbt*^h homozygotes²⁰ (Supplementary Table 1). This shows that the period lengthening we observed in the fly system is indeed a consequence of the mutation and not an artefact caused by overexpressing the *hCKIδ* or *dbt* genes.

We also created transgenic mice with a human BAC clone (RP11-1376P16) containing the entire wild-type CKIδ gene. This

Table 1 Period variation in control and transgenic flies

Genotype	Line	Period length (h)	s.d. (h)	Number of rhythmic flies	Total number tested
<i>tim-UAS-gal4/UAS-hCK1δ</i>	Line 1 (E25)	25.54	0.31	38	39
	Line 2 (E27)	25.39	0.28	27	27
	Line 3 (F5)	25.14	0.32	38	38
<i>tim-UAS-gal4/UAS-hCK1δ-T44A</i>	Line 1 (H3)	25.96	0.39	36	36
	Line 2 (H8)	26.00	0.32	35	36
	Line 3 (H10)	25.86	0.26	24	26
<i>UAS-hCK1δ/+</i>	Line 1 (E25)	23.56	0.29	9	10
	Line 2 (E27)	23.88	0.39	13	13
	Line 3 (F5)	23.80	0.26	9	10
<i>UAS-hCK1δ-T44A/+</i>	Line 1 (H3)	23.68	0.36	10	10
	Line 2 (H8)	23.60	0.46	10	10
	Line 3 (H10)	23.67	0.34	10	10
<i>tim-UAS-gal4/+</i>		24.13	0.28	10	10
<i>w*</i>		23.57	0.30	18	21

Period length (τ) was significantly longer in *tim-UAS-gal4/UAS-hCK1δ-T44A* flies compared with *tim-UAS-gal4/UAS-hCK1δ* flies ($P < 0.0001$, ANOVA). **w* is a recessive allele (*white*) affecting eye pigmentation. It serves as a control because all the transgenic lines were created using this background.

BAC clone was modified to introduce the T44A variant (Supplementary Fig. 2a). The resulting BAC construct contains the promoter as well as an additional 32 kilobases (kb) upstream of the gene, and is expected to contain all *cis*-acting regulatory elements needed to accurately reproduce the endogenous pattern of expression in humans²¹. Transgenic mice had no gross phenotypic or behavioural abnormalities. Analysis of transgene copy number showed that two lines contained 2–3 copies (H lines) and four lines had one copy each (L lines) (Supplementary Fig. 2b). RT-PCR with human- and murine-specific primers was used to compare the transcription levels from human *CKIδ* and murine *Csnk1δ* gene expression (Supplementary Fig. 2c). Only transgenic mice gave PCR products when human-specific primers were used, confirming human *CKIδ* expression in these mice.

Semi-quantitative PCR analysis showed that different transgenic lines had endogenous *Csnk1δ* expression levels similar to wild-type mice, and that the expression level of human *CKIδ* is higher in H lines than in L lines. These results imply that BAC transgene expression levels are related to copy number and independent of the site of integration.

Wheel-running activity of wild-type and *hCKIδ-T44A* transgenic mice was examined by entraining the animals to a schedule of 12 h light–12 h dark (LD) for 7 days before transferring them to constant darkness (DD). Under LD and DD conditions, wild-type and *hCKIδ* transgenic lines showed similar and robust circadian rhythms of activity (Fig. 3a–f). Free-running periods were significantly shorter in transgenic animals compared with wild-type mice (analysis of variance (ANOVA), $P < 0.001$; Fig. 3g). To address the possibility of

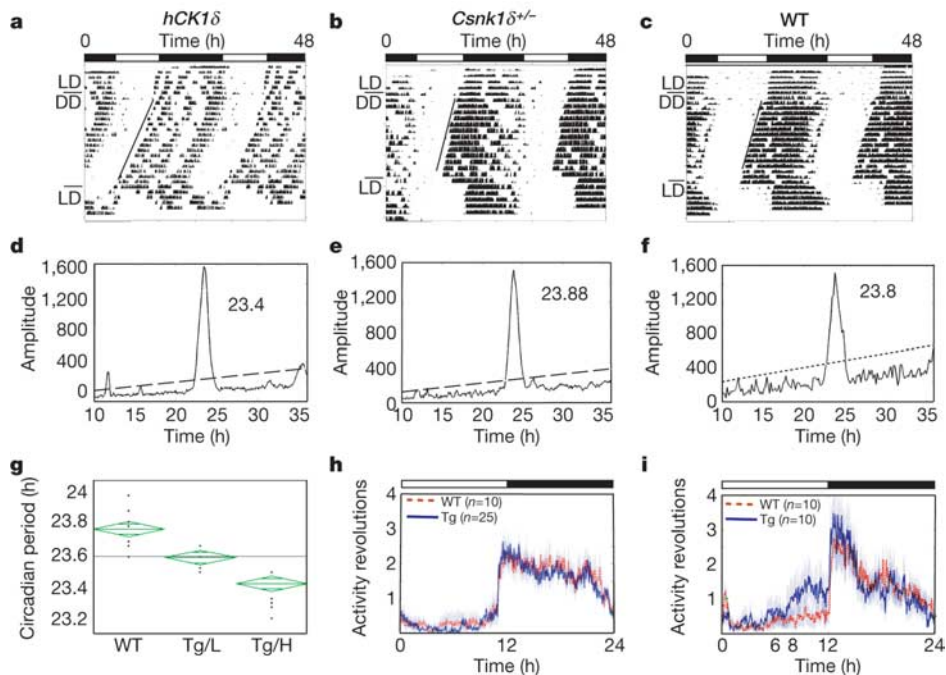


Figure 3 Circadian locomotor activity of *hCKIδ* transgenic mice. **a–c**, Voluntary locomotor activity was recorded as wheel-running activity for *hCKIδ* transgenic mice, *Csnk1δ*^{+/-} heterozygous knockout mice and wild-type mice under LD and DD conditions. **d–f**, Periodograms of locomotor activity for **a–c**, showing periods of 23.4 h (**d**), 23.88 h (**e**) and 23.8 h (**f**). **g**, The distribution of period length determined for days 8–21 in wild-type, low copy number transgenic mice (Tg/L), and high copy number transgenic mice

(Tg/H). Average lengths of the free-running periods ($\pm$ s.d.) were: wild type, 23.78 $\pm$ 0.12 h ($n = 17$); Tg/L, 23.62 $\pm$ 0.07 h ($n = 19$); Tg/H, 23.42 $\pm$ 0.12 h ($n = 17$). $P < 0.001$, ANOVA. **h, i**, Onset activity analysis of wild-type (red) and transgenic (blue) mice during initial LD entrainment (**h**) and re-entrainment after the free-running wheel test (**i**). During entrainment, the light was switched on at $T = 0$ h and switched off at $T = 12$ h.

dosage effect, we generated knockout mice for the human *CKIδ* homologue, *Csnk1δ* (Supplementary Fig. 3). The *Csnk1δ*^{-/-} homozygous mice died within days of birth. We therefore obtained *Csnk1δ*^{+/-} heterozygotes (containing one copy of the wild-type gene), and measured their activity rhythms. These mice showed no change in τ (23.88 ± 0.2 h, $n = 7$) when compared to wild-type mice (23.80 ± 0.1 h, $n = 17$). These results provide strong evidence that the circadian period change in *hCKIδ-T44A* transgenic mice is due to the T44A mutation. These results represent the first direct evidence that *CKIδ* is a core clock component, and that a *CKIδ* mutation gives rise to similar circadian phenotypes in humans and mice.

Activity patterns in wild-type and *hCKIδ-T44A* transgenic mice were indistinguishable under regular LD conditions (Fig. 3h). However, after free running, the *hCKIδ-T44A* transgenic mice required a significantly longer time to re-entrain to a LD regimen compared with wild-type mice (Fig. 3i). This observation is different from the *tau* hamster, for which clear phase-shift under LD conditions was found²². Whether this results from an altered phase angle or from decreased sensitivity to masking by light remains to be elucidated.

Precise control of the mammalian clock requires careful regulation of clock protein production, post-translational modifications and protein–protein interactions^{6,23,24}. Phosphorylation affects the stability and nuclear translocation of clock proteins^{6,11,14,25}. Understanding how CKI mutations alter circadian period length requires an understanding of how the mutations affect the enzymatic properties of CKI and its interactions with clock-relevant substrates. Here, we showed that the human *CKIδ-T44A* mutation has attenuated enzyme activity *in vitro*. This mutation co-segregates with the FASPS phenotype in the affected family but was not found in controls. Whether *CKIδ-T44A* also causes a delay in the appearance of hyper-phosphorylated substrates (as in the *tau* hamster) remains to be seen. Interestingly, the activity pattern of *CKIδ-T44A* mice is different from that of the *tau* hamster even though the mutations are in homologous genes, suggesting that these mutations influence different downstream effects. The results presented here suggest a direct role for *CKIδ* in circadian clock regulation.

We have previously noted a tendency of FASPS-affected individuals to have higher scores on the Beck Depression Inventory¹⁶. It is intriguing that four of the five affected individuals in the K5231 kindred have clinical features or a history of depression. This might be a simple coincidence, or the depression might be situational, resulting from being 'out of phase' with the rest of the world. A more provocative possibility is that circadian rhythm variants contribute to affective behaviour. Sensitivity to light is a shared characteristic of both the circadian system (the process of entrainment) and affective disorders (such as seasonal affective disorder). *shaggy*, a circadian mutant in flies¹¹, results from mutations in *GSK3β*, the protein which is a principal target of lithium, a drug used to treat severe depression²⁶. It is also noteworthy that sleep deprivation is therapeutic in some patients with severe depression that is resistant to pharmacological therapy²⁷. These *CKIδ* transgenic mice will allow testing of this and other possibilities prompted by observations in human FASPS subjects.

It is interesting that the same mutation in the *hCKIδ* gene produces opposite effects on the circadian period when expressed in flies as opposed to mice. This suggests that despite the highly conserved nature of individual components of the circadian clock in mammals and flies, the interactions of these components as part of the circadian regulatory network may be different. The differences in the activity pattern and period length change between transgenic mice carrying the *hCKIδ* mutation and the *tau* hamster indicate that different mutations in the similar protein may have different downstream effects, and that separate regulatory mechanisms are responsible for period length and entrainment. These models

provide unique opportunities to dissect further the molecular mechanisms of circadian clocks, and to perform extensive characterization of a gene variant relevant to human circadian function. New insights into CKI functions could have implications for the development of novel compounds for the treatment of sleep disorders. □

Methods

Diagnostic and classification criteria

Subjects signed a consent form approved by the Institutional Review Board at the University of Utah. Self-reported habitual sleep-wake schedule was obtained during structured interviews by two of the authors (C.R.J., R.E.S.). Self-report was supplemented by parental report for several of the children still at home. Individuals were considered to have FASPS if (1) they habitually preferred to fall asleep before 20:30 and to wake up spontaneously by 05:30 local time throughout the year, in the absence of self-imposed early morning bright lights, stimulant or sedative drugs, early morning work demands, or psychosocial preferences that compete with sleep; (2) the onset of FASPS was before 30 years of age, and not within three months after traumatic brain injury; and (3) there was only one major sleep period per 24-h day¹. Individuals were considered unaffected if they preferred to go to sleep after 21:30 and wake after 06:30 in the absence of psychosocial demands or drug use, and if this preference was stable before the age of 30. Individuals not meeting either 'affected' or 'unaffected' criteria were considered to be of 'unknown' phenotype. Blood sample collection and DNA preparation were performed as previously described²⁸.

In vitro kinase assays

Kinase assays were performed as described elsewhere¹⁴ using 150 ng recombinant *CKIδ*Δ317. Phosvitin, α -casein, or 8 μ l of an *in vitro* translation reaction (Promega) was used as the substrate. For substrates consisting of circadian proteins, kinase was first incubated with translation reaction for 1 h at room temperature, followed by five washes using 1 $\times$ Bind/Wash Buffer (Novagen). Bound beads were then subjected to the kinase assay. The reaction was terminated by the addition of 4 $\times$ LDS sample buffer and 10 $\times$ reducing agent (Invitrogen) followed by a 10 min incubation at 70 °C. Products were resolved on a 12% Bis-Tris NuPAGE gel (Invitrogen). Gels were fixed, dried and exposed to X-OMAT AR film (Kodak). Bands were quantified using NIH Image and substrate-velocity plots were determined using nonlinear regression (GraphPad) to determine V_{max} and K_m .

Fly locomotor activity analysis

Young, male, adult flies of the specified genotype were entrained for 3–4 days in 12 h light–12 h dark (LD) conditions. Locomotor activity was subsequently recorded in constant darkness (DD) for 7–8 days using Trikinetics locomotor activity monitors at 25 °C. The circadian period of the flies was calculated using χ^2 periodogram analysis ($\alpha = 0.01$) on the ClockLab software package. Only flies for which the χ^2 statistic exceeded the significance line by 5 were judged to be rhythmic and used for period calculations.

Generation of mice transgenic for human *CKIδ*

Transgenic mice were generated using standard procedures. The transgenic founders were on a C57BL/6 $\times$ SJL F₁ background and were backcrossed to C57BL/6 mice in successive generations. Founders were determined by PCR analysis using 6 primer pairs every 15 kb around this gene. Subsequent to the first generation, genotyping was performed using PCR analysis of tail biopsy DNA with primers (5'-AAGAGGGGAGTACGTGTAGAAC-3' and 5'-AATATAAATGTGGGGACTGAGCAT-3'). The PCR protocol consisted of 3 min at 94 °C, 35 cycles of amplification (30 s at 94 °C, 30 s at 57 °C, and 90 s at 72 °C), and a final extension phase (10 min at 72 °C), and yielded a ~500 bp band. Transgene copy number was estimated by Southern blot. Tail DNA or 0X, 1X, 3X, 5X and 10X BAC DNA-spiked tail DNA (10 μ g per lane) was digested with *Bgl*III, electrophoresed on 0.8% agarose, and transferred to a Hybond N+ membrane. The blot was hybridized with a random prime-³²P-dCTP-labelled probe amplified from the BAC clone. Inclusion of the copy standards allowed us to calculate the approximate number of transgenes integrated for each line. For RT-PCR, total RNA was extracted from mouse brain with TRIzol (Invitrogen) according to the manufacturer's instructions, then subjected to DNase treatment (Ambion). RNA (10 μ g) was reverse-transcribed using Superscript III (Invitrogen). The primer sequences for human *CKIδ* were 5'-CTGCTCCAGGAAATTCAGCC-3' and 5'-AGGTCGGACGAGGAGATGTT-3'. The primer sequences for mouse *Csnk1δ* were 5'-TCTGTTCAAGGAAGTTAGT-3' and 5'-AGATCAGATGAGGAGACGTT-3'. The primer sequences for the actin gene (used as an endogenous control reference) were 5'-CTCTTTGATGTCACGACAGATTTC-3' and 5'-GTGGGCCGCTCTAGGCACCAA-3'. The following conditions were used for PCR reactions: 3 min at 94 °C, 30 cycles of amplification (30 s at 94 °C, 30 s at 60 °C, and 90 s at 72 °C), followed by 7 min final extension at 72 °C.

Mouse behavioural analysis

Three- to four-month-old animals were housed individually in cages equipped with running wheels, and exposed to a 12 h light–12 h dark cycle for one week before being released into constant darkness (wild type, $n = 17$; *CKIδ* transgenic mice, $n = 37$). Dim red light was present in each compartment at all times, including those designated as 'dark'. Activity data were collected^{29,30}. Using ClockLab (Actimetrics) software, data

analysis included a χ^2 periodogram, fast Fourier transformation (FFT), least squares fit of activity onset, and total activity analyses. The χ^2 periodogram was performed at 1-min resolution for days 1–14 of constant darkness, and saved in digital format.

Received 9 November 2004; accepted 4 February 2005; doi:10.1038/nature03453.

1. Jones, C. R. *et al.* Familial advanced sleep-phase syndrome: A short-period circadian rhythm variant in humans. *Nature Med.* **5**, 1062–1065 (1999).
2. Dunlap, J. C. Molecular bases for circadian clocks. *Cell* **96**, 271–290 (1999).
3. Edery, I., Zwiebel, L. J., Dembinska, M. E. & Rosbash, M. Temporal phosphorylation of the *Drosophila* period protein. *Proc. Natl Acad. Sci. USA* **91**, 2260–2264 (1994).
4. Denault, D. L., Loros, J. J. & Dunlap, J. C. WC-2 mediates WC-1–FRQ interaction within the PAS protein-linked circadian feedback loop of *Neurospora*. *EMBO J.* **20**, 109–117 (2001).
5. Young, M. W. Life's 24-hour clock: molecular control of circadian rhythms in animal cells. *Trends Biochem. Sci.* **25**, 601–606 (2000).
6. Lee, C., Etchegaray, J. P., Cagampang, F. R., Loudon, A. S. & Reppert, S. M. Posttranslational mechanisms regulate the mammalian circadian clock. *Cell* **107**, 855–867 (2001).
7. Nawathean, P. & Rosbash, M. The doubletime and CKII kinases collaborate to potentiate *Drosophila* PER transcriptional repressor activity. *Mol. Cell* **13**, 213–223 (2004).
8. Sathyanarayanan, S., Zheng, X., Xiao, R. & Sehgal, A. Posttranslational regulation of *Drosophila* PERIOD protein by protein phosphatase 2A. *Cell* **116**, 603–615 (2004).
9. Kloss, B. *et al.* The *Drosophila* clock gene double-time encodes a protein closely related to human casein kinase I ϵ . *Cell* **94**, 97–107 (1998).
10. Price, J. L. *et al.* double-time is a novel *Drosophila* clock gene that regulates Period protein accumulation. *Cell* **94**, 83–95 (1998).
11. Martinek, S., Inonog, S., Manoukian, A. S. & Young, M. W. A role for the segment polarity gene *shaggy*/GSK-3 in the *Drosophila* circadian clock. *Cell* **105**, 769–779 (2001).
12. Lin, J. M. *et al.* A role for casein kinase 2 α in the *Drosophila* circadian clock. *Nature* **420**, 816–820 (2002).
13. Akten, B. *et al.* A role for CK2 in the *Drosophila* circadian oscillator. *Nature Neurosci.* **6**, 251–257 (2003).
14. Lowrey, P. L. *et al.* Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau. *Science* **288**, 483–492 (2000).
15. Preuss, F. *et al.* *Drosophila* doubletime mutations which either shorten or lengthen the period of circadian rhythms decrease the protein kinase activity of casein kinase I. *Mol. Cell Biol.* **24**, 886–898 (2004).
16. Beck, A. T. *The Beck Depression Inventory* (Harcourt Brace Jovanich, The Psychological Corporation, San Antonio, 1978).
17. Dahl, R. E. *et al.* Sleep onset abnormalities in depressed adolescents. *Biol. Psychiatry* **39**, 400–410 (1996).
18. Graves, P. R., Haas, D. W., Hagedorn, C. H., DePaoli-Roach, A. A. & Roach, P. J. Molecular cloning, expression, and characterization of a 49-kilodalton casein kinase I isoform from rat testis. *J. Biol. Chem.* **268**, 6394–6401 (1993).
19. Blau, J. & Young, M. W. Cycling vrille expression is required for a functional *Drosophila* clock. *Cell* **99**, 661–671 (1999).
20. Suri, V., Hall, J. C. & Rosbash, M. Two novel doubletime mutants alter circadian properties and eliminate the delay between RNA and protein in *Drosophila*. *J. Neurosci.* **20**, 7547–7555 (2000).
21. Heintz, N. BAC to the future: the use of bac transgenic mice for neuroscience research. *Nature Rev. Neurosci.* **2**, 861–870 (2001).
22. Ralph, M. R. & Menaker, M. A mutation of the circadian system in golden hamsters. *Science* **241**, 1225–1227 (1988).
23. Shearman, L. P. *et al.* Interacting molecular loops in the mammalian circadian clock. *Science* **288**, 1013–1019 (2000).
24. Lee, C., Weaver, D. R. & Reppert, S. M. Direct association between mouse PERIOD and CKI ϵ is critical for a functioning circadian clock. *Mol. Cell Biol.* **24**, 584–594 (2004).
25. Eide, E. J., Vielhaber, E. L., Hinz, W. A. & Virshup, D. M. The circadian regulatory proteins BMAL1 and cryptochromes are substrates of casein kinase I ϵ . *J. Biol. Chem.* **277**, 17248–17254 (2002).
26. Phiel, C. J. & Klein, P. S. Molecular targets of lithium action. *Annu. Rev. Pharmacol. Toxicol.* **41**, 789–813 (2001).
27. Gillin, J. C. The sleep therapies of depression. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **7**, 351–364 (1983).
28. Toh, K. L. *et al.* An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome. *Science* **291**, 1040–1043 (2001).
29. Vitaterna, M. H. *et al.* Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. *Science* **264**, 719–725 (1994).
30. Antoch, M. P. *et al.* Functional identification of the mouse circadian Clock gene by transgenic BAC rescue. *Cell* **89**, 655–667 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The authors thank the FASPS subjects and their families for participating in this research. We thank U. Heberlein for advice and generous use of fly laboratory facilities, R. Threlkeld for *Drosophila* injections, and A. Rothenfluh for discussions and support with *Drosophila* transgenic lines. We also thank M. W. Young and L. Saez for the *tim-UAS-gal4* stock and helpful discussions, and S. Reppert for the mouse *Per1* clone. We acknowledge J. Cheung, E. Stryker and C. Whitney for technical assistance and members of the Fu and Ptáček laboratories for discussions. R.E.S. is supported by an NIH GCRC grant and the FAHC/UMV Office of Patient Oriented Research. This work was supported by an NIH grant to Y.-H.F. and L.J.P., and a Sandler Neurogenetics grant to Y.-H.F. L.J.P. is an investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.-H.F. (yinghui@itsa.ucsf.edu).

Deletion of active ADAMTS5 prevents cartilage degradation in a murine model of osteoarthritis

Sonya S. Glasson¹, Roger Askew², Barbara Sheppard³, Brenda Carito³, Tracey Blanchet¹, Hak-Ling Ma¹, Carl R. Flannery¹, Diane Peluso¹, Kim Kanki², Zhiyong Yang¹, Manas K. Majumdar¹ & Elisabeth A. Morris¹

¹Department of Women's Health and Bone, ²Genomics and ³Drug Safety and Metabolism, Wyeth Research, 200 CambridgePark Drive, Cambridge, Massachusetts 02140, USA

Human osteoarthritis is a progressive disease of the joints characterized by degradation of articular cartilage. Although disease initiation may be multifactorial, the cartilage destruction appears to be a result of uncontrolled proteolytic extracellular matrix destruction. A major component of the cartilage extracellular matrix is aggrecan, a proteoglycan that imparts compressive resistance to the tissue. Aggrecan is cleaved at a specific 'aggrecanase' site in human osteoarthritic cartilage^{1–2}; this cleavage can be performed by several members of ADAMTS family of metalloproteases^{3–9}. The relative contribution of individual ADAMTS proteases to cartilage destruction during osteoarthritis has not been resolved. Here we describe experiments with a genetically modified mouse in which the catalytic domain of ADAMTS5 (aggrecanase-2) was deleted. After surgically induced joint instability, there was significant reduction in the severity of cartilage destruction in the ADAMTS5 knockout mice compared with wild-type mice. This is the first report of a single gene deletion capable of abrogating the course of cartilage destruction in an animal model of osteoarthritis. These results demonstrate that ADAMTS5 is the primary 'aggrecanase' responsible for aggrecan degradation in a murine model of osteoarthritis, and suggest rational strategies for therapeutic intervention in osteoarthritis.

The two major structural components of cartilage extracellular matrix are the proteoglycan aggrecan and type II collagen. Pathologic cleavage of aggrecan at Glu 373/Ala 374 (the 'aggrecanase' site) was identified as the major site of aggrecan degradation in human joint disease by analysis of synovial fluid samples from a range of human joint pathologies including osteoarthritis^{1,2}. This is also the primary site of aggrecan cleavage in response to inflammatory stimuli¹⁰. Several members of the ADAMTS (a disintegrin and metalloprotease with thrombospondin-like repeat) family of enzymes (ADAMTS1, 4, 5, 8, 9 and 15) are known to be capable of cleaving aggrecan at the Glu 373/Ala 374 site^{3–9}, but ADAMTS4 and ADAMTS5 (aggrecanase-1 and aggrecanase-2, respectively) seem to be the most active aggrecanases^{8,9}. Which aggrecanase is responsible for aggrecan degradation during human articular cartilage destruction, however, remains unclear. We addressed this question using gene-targeted deletion of the catalytic domains of ADAMTS4 or ADAMTS5 in mice, followed by induction of cartilage degradation by surgically induced joint instability. The functions of ADAMTS4 and 5 activity in normal development and physiology were also investigated by analysis of the histological appearance of organs in adult animals.

We have previously reported the generation of ADAMTS4 knockout (ADAMTS4^{-/-}) mice¹¹. The ADAMTS5 (ADAMTS5^{-/-}) knockout was created by cre/lox mediated recombination, resulting in the deletion of 56 amino acids encoded by exon 3, including disruption of the zinc-binding site and deletion of the conserved 'Met-turn' (Fig. 1a). The deletion was confirmed by polymerase chain reaction (PCR) from tail DNA (Fig. 1b). PCR with reverse transcription (RT-PCR) from spleen total RNA gave PCR products

consistent with ADAMTS5 messenger RNA lacking exon 3 (Fig. 1c). The presence of ADAMTS5 mRNA was confirmed in the articular cartilage (Fig. 1d), but the presence of the translated protein lacking the catalytic domain could not be confirmed in either wild-type or ADAMTS5^{-/-} articular cartilage. (See Supplementary Methods and Results.)

Adult (18-week-old) male and female wild-type and homozygous knockout mice were examined for gross abnormalities. Blood samples were drawn for complete blood count and serum chemistry, and 17 tissues were harvested and examined by a board-certified veterinary pathologist. Tissues examined included heart, lung, thymus, spleen, liver, kidney, brain, salivary glands, mandibular lymph node, testis, epididymis, eye, harderian gland, lacrimal gland, whole sternum, femur and whole paw. There were no abnormalities in total body weight, blood or serum analysis or histological appearance of any tissue examined, indicating that ADAMTS5 enzyme activity is not critical for normal development and growth.

Aggrecan is an abundant component of cartilage within the growth plate as well as the articular cartilage of the joints. Analysis of growth plates from wild-type, ADAMTS4^{-/-} and ADAMTS5^{-/-} animals was performed by immunostaining the proximal tibial growth plates with an antibody raised against a C-terminal neopeptide of aggrecan, G1-TEGE³⁷³ (where the super-

script number denotes the residue position in aggrecan), generated following cleavage by aggrecanases. Significant staining was observed in the cells of the proximal tibial growth plate in wild-type mice, but staining was negligible in the ADAMTS4^{-/-} animals¹¹. In contrast, the staining of G1-TEGE³⁷³ within the growth plate of the ADAMTS5^{-/-} mice was similar to that in wild-type growth plates (Fig. 2). These results suggest that aggrecanase-mediated aggrecan turnover in the murine growth plate is a result of ADAMTS4 activity and not ADAMTS5. The gross appearance of the animals, the length of long bones and the histological appearance of the growth plates in both knockout strains were identical to the wild-type mice. It therefore appears that the aggrecan-degrading activity of ADAMTS4 in the growth plate is not essential for normal growth.

Unilateral joint instability was generated by surgical transection of the medial meniscotibial ligament. Mice were sacrificed 4 and 8 weeks after surgery, and sections through the joint were scored using a previously reported scoring system^{11,12}. Scores were reported as mean maximal scores from wild-type, ADAMTS5^{+/-} and ADAMTS5^{-/-} mice. In addition, the scores from each section through the joint were added together, to express the severity of pathology as the summed score for each treatment group at both 4 and 8 weeks after surgery. This second method of scoring takes into

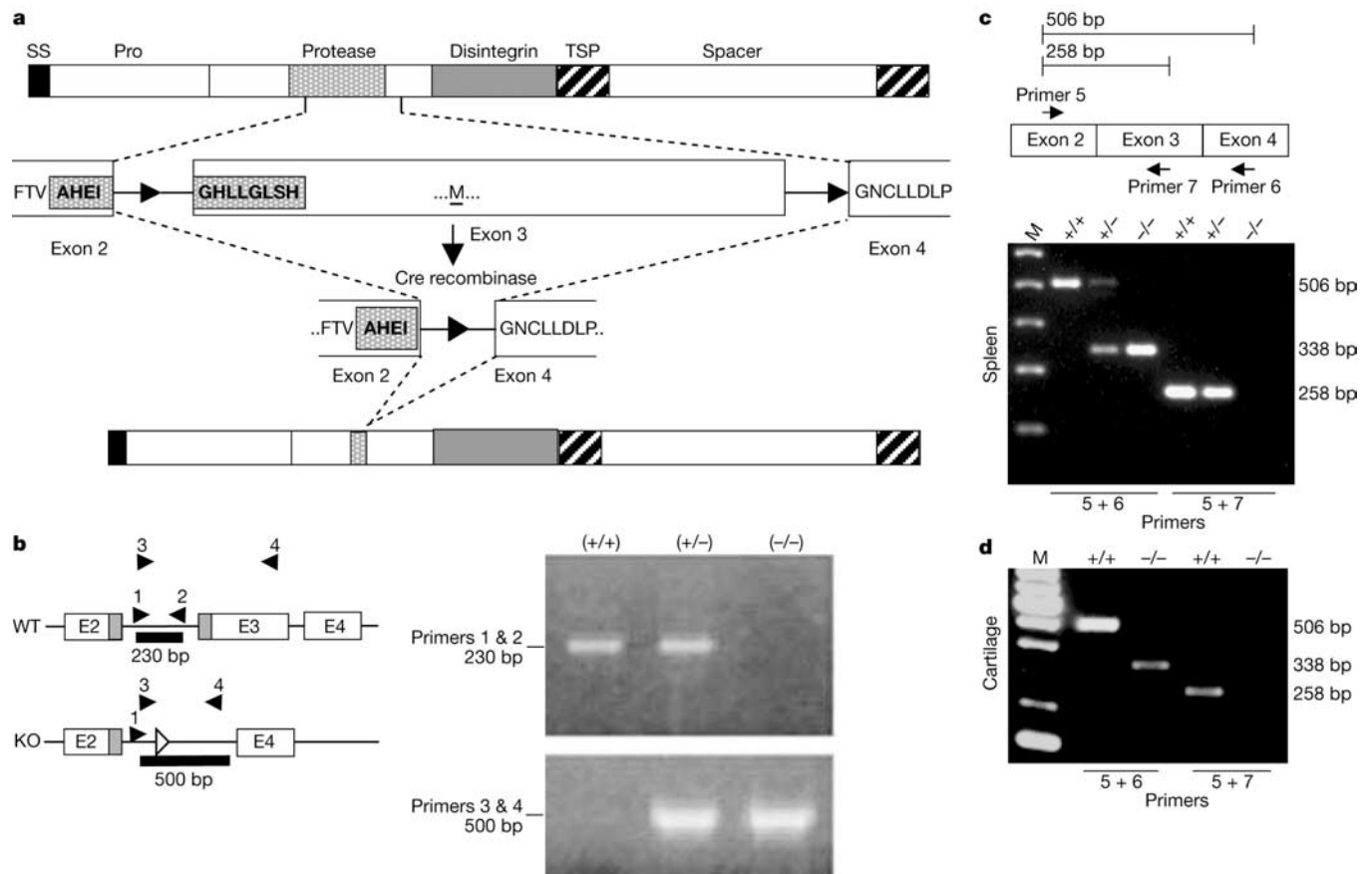


Figure 1 Targeted disruption of the murine ADAMTS5 gene. **a**, Diagram of the ADAMTS5 protein and the strategy used for creation of the knockout. The protease consensus sequence HEXXH (shaded box) is divided between exon 2 and exon 3. The invariant 'Met-turn' in exon 3 is underlined. LoxP sites are shown by arrows. SS, signal sequence; TSP, thrombospondin-like domain. **b**, Allele characterization of the ADAMTS5 knockout mice. The protease domain is depicted as a shaded box. PCR products from wild-type (+/+) and heterozygous (+/-) mice (using primers 1 and 2) generated a 230 base pair (bp) product. Primer 2 was located within the deleted portion and therefore did not generate a PCR product in the homozygous (-/-) knockout mice. Primers 3 and 4

generated a 500 bp product in ADAMTS^{+/-} and ADAMTS^{-/-} mice owing to deletion of exon 3. **c**, **d**, ADAMTS5 mRNA from spleen (**c**) and cartilage (**d**). RT-PCR using primers 5 and 6 generated a 506 bp product in the wild-type animals and a 338 bp product in the knockouts, indicating deletion of exon 3. Both fragments were amplified in the heterozygous mice. RT-PCR using primers 5 and 7 generated a 258 bp product in both wild-type and heterozygous animals, and no PCR product in the homozygous knockout because primer 7 was designed to recognize a site within the deleted exon 3. M indicates base pair markers.

account severity of the lesion as well as the surface area of the joint affected. Both methods of analysis revealed a significant reduction ($P < 0.05$) in the scores of the ADAMTS^{+/−} and ADAMTS^{−/−} mice compared with wild-type mice (Fig. 3, see Supplementary Figs 1 and 2). The summed scores of the ADAMTS5 knockout animals were less than 50% of the wild type, as a result of reduced severity and surface area involvement. The severity of disease in heterozygous knockout animals was also significantly reduced 8 weeks after induction of joint instability.

This model of cartilage degradation is characterized by initial aggrecan depletion of intact cartilage followed by fibrillation, clefting and subsequent loss of cartilage down to the tidemark¹². Using the anti-G1-TEGE³⁷³ antibody to immunostain joints from wild-type mice 8 weeks after induction of joint instability revealed significant amounts of cleaved aggrecan in the intact articular cartilage; this cleavage product was markedly reduced in the corresponding intact ADAMTS^{−/−} cartilage (see Supplementary Information). This observation supports the importance of ADAMTS5 in the early phase of the degradative process. No cartilage protection was observed in ADAMTS4^{−/−} mice using the identical surgical model¹¹. Similar studies describing induction of osteoarthritis in mice with deletions of interleukin-1 β (IL-1 β), inducible nitric oxide synthase (iNOS), Stromelysin 1 or IL-1 β converting enzyme (ICE) have reported increased severity of

pathology in the genetically manipulated mice¹³.

Femoral head articular cartilage was removed from ADAMTS^{−/−}, ADAMTS^{+/−} and wild-type mice and placed into tissue culture¹⁴. Inflammatory modulators (IL-1 α and retinoic acid) were added to the culture system to induce degradative enzyme activity. A significant increase in total proteoglycan and G1-TEGE³⁷³ release by wild-type cartilage was observed after exposure to IL-1 and retinoic acid, but minimal evidence was found for proteoglycan release by ADAMTS^{−/−} cartilage in a similar inflammatory environment (Fig. 4). This is in marked contrast with the abundant generation of the TEGE³⁷³ fragment in cultured cartilage from ADAMTS4^{−/−} mice, at comparable levels to wild type¹¹. Immunohistochemical analysis of these femoral head cartilage explants using an anti-TEGE³⁷³ antibody showed abundant intracellular and pericellular TEGE³⁷³ in articular cartilage from wild-type and ADAMTS4^{−/−} mice, but this cleavage product was essentially undetectable in ADAMTS^{−/−} cartilage (Fig. 4c).

This study examined the effect of deleting ADAMTS5 activity on normal murine development, growth and physiology and on the progression of instability-induced cartilage destruction in the murine stifle. Similar to our previous report examining the ADAMTS4^{−/−} mouse¹¹, deletion of ADAMTS5 activity did not negatively affect normal development or growth. Deletion of ADAMTS5 activity did not negatively affect aggrecanase-mediated aggrecan turnover in the growth plates. However, deletion of ADAMTS5 activity significantly abrogated the progression of cartilage degradation in mice. This inability to cleave aggrecan at the aggrecanase site (located in the interglobular domain of the substrate) was further substantiated by lack of appearance of these fragments after *in vitro* cytokine stimulation of articular cartilage from ADAMTS^{−/−} mice. Although several aggrecanases have been reported to be capable of cleaving aggrecan at the Glu 373/Ala 374 site, the present data suggest that ADAMTS5 is the principal enzyme

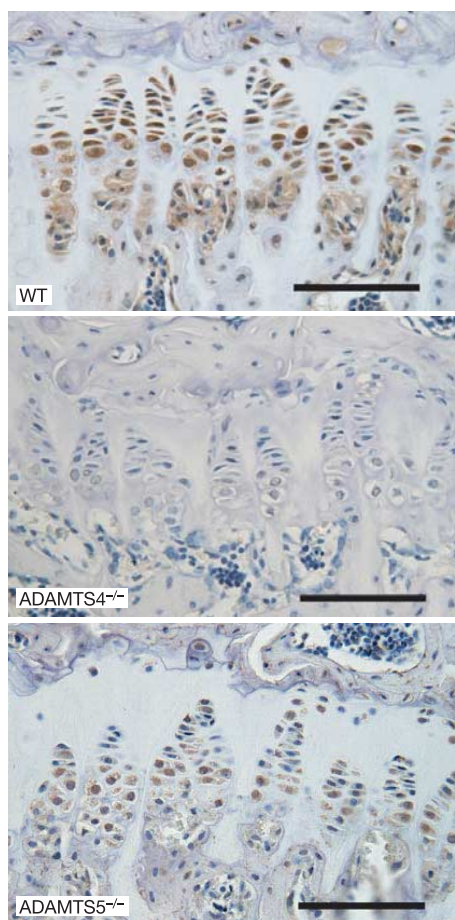


Figure 2 Immunohistochemical localization of the aggrecanase-generated TEGE³⁷³ neopeptide in growth plates of wild-type, ADAMTS4^{−/−} and ADAMTS5^{−/−} mice. Cleavage of aggrecan at the TEGE³⁷³ site is visible (as brown staining) in the wild-type and ADAMTS5^{−/−} growth plates, but is negligible in the ADAMTS4^{−/−} mice, indicating that ADAMTS4 is the sole aggrecanase responsible for aggrecan cleavage at this anatomical location. The histological appearance of growth plates from all 3 mice is normal. Original magnification $\times 20$; scale bar, 100 μ m.

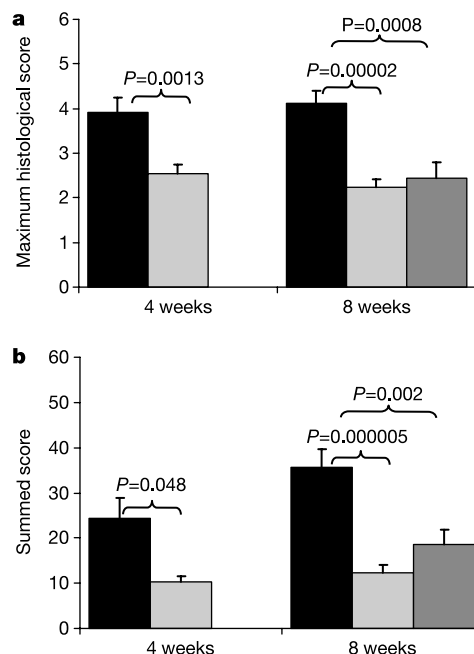


Figure 3 Histological scores of joints 4 and 8 weeks after induction of joint instability. Wild-type (black), ADAMTS4^{−/−} (dark grey) and ADAMTS5^{−/−} (light grey) mice. Histological scores are expressed as the mean maximal score from each joint (a), or as the mean of the sum of the scores from each histological section through the joints (b). The latter method of scoring takes into account both severity of the lesions as well as the surface area of the joint demonstrating pathology. Error bars indicate standard error of the mean.

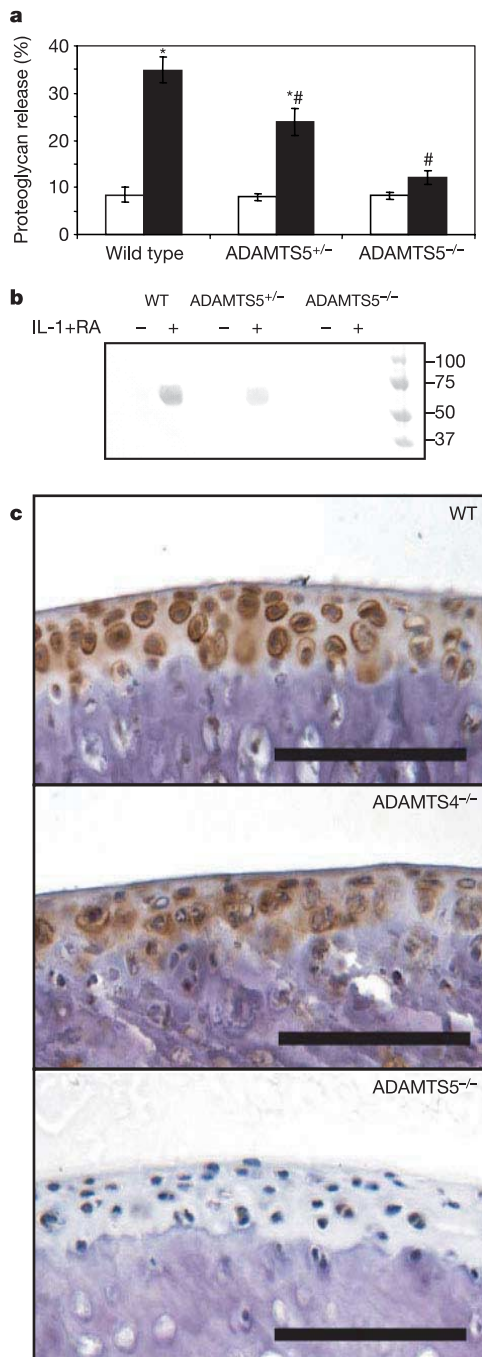


Figure 4 Proteoglycan release from articular cartilage from wild-type and ADAMTS5 knockout mice. **a**, Percentage of total proteoglycan released from cultured articular cartilage in medium alone (open bars) or in medium containing IL-1 α and retinoic acid (RA) (filled bars). Single asterisk denotes a significant difference ($P < 0.05$) from wild-type cartilage in medium alone; single hash denotes a significant difference ($P < 0.05$) from wild-type cartilage in medium + IL-1 + RA. Error bars indicate s.e.m. **b**, Western blot analysis of the TEGE³⁷³ fragment released from articular cartilage in the presence (+) or absence (-) of IL-1 and RA. The lack of release of the TEGE³⁷³ fragment from ADAMTS5^{-/-} articular cartilage implicates ADAMTS5 as the source of the aggrecanase degradation products. **c**, TEGE³⁷³ immunostaining (brown staining) of articular cartilage from wild-type (WT), ADAMTS4^{-/-} and ADAMTS5^{-/-} articular cartilage in the presence of IL-1 and RA. Absence of the TEGE³⁷³ fragment in the ADAMTS5^{-/-} articular cartilage indicates no aggrecanase activity. Original magnification $\times 20$; scale bar, 100 μm .

responsible for articular cartilage aggrecan degradation in this murine model of osteoarthritis. Corroboration of these findings in human osteoarthritis will require examination of human osteoarthritic cartilage and demonstration of therapeutic efficacy of targeted inhibitors in the clinic. □

Methods

Generation of ADAMTS5 knockout mice

ADAMTS5 knockout mice used in these studies were designed at Wyeth Research and generated from 129SvEv-Brd mice carrying a cre/loxP conditional knockout allele of the *Adamts5* gene (Lexicon Genetics, Inc.). The conditional allele contained loxP sites flanking exon 3 so that cre recombination resulted in deletion of exon 3, which encodes the majority of the enzyme active site.

Genotyping for wild-type and ADAMTS5 knockout alleles

Genotyping for alleles of the *Adamts5* gene was performed by PCR using DNA template from proteinase K digests of tail biopsies. The wild-type and heterozygous knockout alleles were identified by PCR using forward primer 1 (5'-TGTTCA CCCAAAGCACTAC-3') and reverse primer 2 (5'-TAGAGGAGAGGAGAGGAGG-3'). The homozygous knockout allele was identified using forward primer 3 (5'-GTGAACCATGGACTTTGG-3') and reverse primer 4 (5'-TCGTAGCAAACACCCACCTG-3').

Characterization of ADAMTS5 mRNA

Total RNA was prepared from the spleen and cartilage of wild-type and knockout mice using an RNeasy kit (Qiagen). PCR primers were designed and synthesized (BioSource International) as follows: Primer 5 (exon2), 5'-CGACCCCTCAAGAACTTTTGC-3'; Primer 6 (exon 4), 5'-CTCAGGCCCAATGTCAAGT-3'; Primer 7 (exon 3), 5'-CGTCATGAGAAAGGCCAAGT-3'.

Histology and pathological assessment of mice

Complete necropsies were performed on 22 wild-type and 20 ADAMTS5^{-/-} mice at 18 weeks of age. Necropsies included macroscopic observations, body and organ weight measurements, haematology, complete serum chemistry analysis and microscopic examination of 17 tissues (see text). All slides were evaluated by a board-certified veterinary pathologist for the presence or absence of pathological lesions.

Surgical induction of osteoarthritis

All studies were performed with approval of the Wyeth Institutional Animal Care and Use committee. Ten-week-old mice were anaesthetized and microsurgery was performed to transect the meniscotibial ligament (anchoring the medial meniscus to the tibial plateau), resulting in destabilization of the medial meniscus. Four weeks after surgery, 19 wild-type and 19 ADAMTS5^{-/-} mice were sacrificed. Eight weeks after surgery, 33 wild-type, 20 ADAMTS5^{+/-} and 28 ADAMTS5^{-/-} mice were sacrificed.

Assessment of progression and severity of osteoarthritis

Intact knee joints were fixed, decalcified and embedded in paraffin, and frontal sections were taken through the entire joint. Slides were stained with Safranin-O/Fast green and graded at 70- μm intervals through the joint by three scorers who were blind to the specimen samples. The semiquantitative scoring system was modified from previous reports^{11,12}, and graded each quadrant of the joint with a score from 0 (normal cartilage) through 6 (>80% loss of non-calcified cartilage). The medial tibial plateau, medial femoral condyle, lateral tibial plateau and lateral femoral condyle were scored separately. Scores were expressed as the 'maximum histological score' found at all sections through the entire joint. Scores were added from all levels of all four quadrants to obtain the 'summed histological score', which reflects osteoarthritic lesion severity as well as the surface area affected.

Culture of cartilage explants

Femoral head cartilage was harvested from 3–4-week-old wild-type and knockout mice, and cultured in serum-free DMEM medium plus 10 ng ml⁻¹ mouse IL-1 α (Sigma) and 10⁻⁵ M retinoic acid (Sigma) as described¹⁴. After three days in culture, the proteoglycan content in the medium and digested cartilage were measured using a previously reported procedure¹⁵.

Western analysis

Aggrecan fragments in conditioned medium were analysed by western blot as previously described¹⁶. Immunoblotting was performed using monoclonal antibody AGG-C1 (0.04 $\mu\text{g ml}^{-1}$, ref. 6).

Identification of TEGE³⁷³ in murine articular cartilage

Cartilage tissues were frozen in Optimal Cutting Temperature Medium (Tissuetek) and 5- μm sections were cut and immunostained as described¹¹ using a polyclonal anti-TEGE³⁷³ antibody. Immunohistochemistry of the growth plates was performed similarly with the exception that decalcified, paraffin embedded sections were used.

Received 30 November 2004; accepted 19 January 2005; doi:10.1038/nature03369.

1. Sandy, J. D., Flannery, C. R., Neame, P. J. & Lohmander, L. S. The structure of aggrecan fragments in human synovial fluid. Evidence for the involvement in osteoarthritis of a novel proteinase which cleaves the Glu373-Ala374 bond of the interglobular domain. *J. Clin. Invest.* **89**, 1512–1516 (1992).
2. Lohmander, L. S., Neame, P. J. & Sandy, J. D. The structure of aggrecan fragments in human synovial

- fluid. Evidence that aggrecanase mediates cartilage degradation in inflammatory joint disease, joint injury, and osteoarthritis. *Arthritis Rheum.* **36**, 1214–1222 (1993).
3. Kuno, K. *et al.* ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan. *FEBS Lett.* **478**, 241–245 (2000).
 4. Rodriguez-Manzanique, J. C. *et al.* ADAMTS1 cleaves aggrecan at multiple sites and is differentially inhibited by metalloproteinase inhibitors. *Biochem. Biophys. Res. Commun.* **293**, 501–508 (2002).
 5. Somerville, R. P. T. *et al.* Characterization of ADAMTS-9 and ADAMTS-20 as a distinct ADAMTS subfamily related to *Caenorhabditis elegans* GON-1. *J. Biol. Chem.* **278**, 9503–9513 (2003).
 6. Collins-Racie, L. A. *et al.* ADAMTS-8 exhibits aggrecanase activity and is expressed in human articular cartilage. *Matrix Biol.* **23**, 219–230 (2004).
 7. Yamaji, N. *et al.* Metalloprotease having aggrecanase activity. US patent 6,716,613 (2004).
 8. Tortorella, M. D. *et al.* Purification and cloning of aggrecanase-1: a member of the ADAMTS family of proteins. *Science* **284**, 1664–1666 (1999).
 9. Abbaszade, I. *et al.* Cloning and characterization of ADAMTS11, an aggrecanase from the ADAMTS family. *J. Biol. Chem.* **274**, 23443–23450 (1999).
 10. Sandy, J. D., Neame, P. J., Boynton, R. E. & Flannery, C. R. Catabolism of aggrecan in cartilage explants. Identification of a major cleavage site within the interglobular domain. *J. Biol. Chem.* **266**, 8683–8685 (1991).
 11. Glasson, S. S. *et al.* Characterization of and osteoarthritis susceptibility in ADAMTS-4-knock out mice. *Arthritis Rheum.* **50**, 2547–2558 (2004).
 12. Chambers, M. G. *et al.* Matrix metalloproteinases and aggrecanases cleave aggrecan in different zones of normal cartilage but colocalize in the development of osteoarthritic lesions in STR/ort mice. *Arthritis Rheum.* **44**, 1455–1465 (2001).
 13. Clements, K. M. *et al.* Gene deletion of either interleukin-1 β , interleukin-1 β -converting enzyme, inducible nitric oxide synthase or stromelysin 1 accelerates the development of knee osteoarthritis in mice after surgical transection of the medial collateral ligament and partial medial meniscectomy. *Arthritis Rheum.* **48**, 3452–3463 (2003).
 14. Little, C. B., Meeker, C. T. & Fosang, A. J. Degradative mechanisms in mouse articular cartilage: use of *in vitro* models to analyse proteolysis and loss of aggrecan and link protein. *Trans. Orthop. Res. Soc.* **28**, 699 (2003).
 15. Farndale, R. W., Buttle, D. J. & Barrett, A. J. Improved quantitation and discrimination of sulphated glycosaminoglycans by use of dimethylmethylene blue. *Biochim. Biophys. Acta* **883**, 173–177 (1986).
 16. Hughes, C. E., Little, C. B. & Caterson, B. Measurement of aggrecanase-generated interglobular domain catabolites in the medium and extracts of cartilage explants using Western blot analysis. *Methods Mol. Biol.* **225**, 89–98 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The authors would like to acknowledge M. Jin and X.-Y. Tan for generation of the monoclonal antibody AGG-C1. We also acknowledge the assistance of W. Kuang, M. Liang, K. Shih, J. Tavares, K. Tomkinson and T. Crawford, and the Bioresources staff at Wyeth.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.A.M. (Emorris@Wyeth.com).

ADAMTS5 is the major aggrecanase in mouse cartilage *in vivo* and *in vitro*

Heather Stanton¹, Fraser M. Rogerson¹, Charlotte J. East¹, Suzanne B. Golub¹, Kate E. Lawlor¹, Clare T. Meeker¹, Christopher B. Little², Karena Last¹, Pamela J. Farmer³, Ian K. Campbell⁴, Anne M. Fourie⁵ & Amanda J. Fosang¹

¹University of Melbourne Department of Paediatrics and Murdoch Childrens Research Institute, Royal Children's Hospital, Parkville, 3052 Victoria, Australia

²Raymond Purves Bone and Joint Research Laboratories, University of Sydney, Royal North Shore Hospital, St. Leonards, 2065 New South Wales, Australia

³Murdoch Childrens Research Institute, Department of Surgery, Royal Children's Hospital, Parkville, 3052 Victoria, Australia

⁴Reid Rheumatology Laboratory, The Walter and Eliza Hall Institute of Medical Research, Parkville, 3052 Victoria, Australia

⁵Johnson & Johnson Pharmaceutical Research and Development, La Jolla, San Diego, California 92121, USA

Aggrecan is the major proteoglycan in cartilage, endowing this tissue with the unique capacity to bear load and resist compression. In arthritic cartilage, aggrecan is degraded by one or more 'aggrecanases' from the ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs¹) family of proteinases. ADAMTS1, 8 and 9 have weak aggrecan-degrading

activity^{2–5}. However, they are not thought to be the primary aggrecanases because ADAMTS1 null mice are not protected from experimental arthritis⁶, and cleavage by ADAMTS8 and 9 is highly inefficient. Although ADAMTS4 and 5 are expressed in joint tissues^{7–13}, and are known to be efficient aggrecanases *in vitro*, the exact contribution of these two enzymes to cartilage pathology is unknown. Here we show that ADAMTS5 is the major aggrecanase in mouse cartilage, both *in vitro* and in a mouse model of inflammatory arthritis. Our data suggest that ADAMTS5 may be a suitable target for the development of new drugs designed to inhibit cartilage destruction in arthritis, although further work will be required to determine whether ADAMTS5 is also the major aggrecanase in human arthritis.

Mice bearing *loxP* sites flanking the exons encoding the catalytic site of ADAMTS4 (exon 4) or ADAMTS5 (exon 3) were a gift from Johnson & Johnson Pharmaceutical Research and Development, and were developed by Lexicon Genetics Incorporated. The targeting constructs, together with the wild-type and modified alleles for *Adamts4* and *Adamts5*, are shown in Fig. 1a, b. To delete the *loxP*-flanked (floxed) region of the gene, homozygous *loxP/loxP* mice were bred with mice expressing the *Cre* transgene under the control of the cytomegalovirus promoter¹⁴. This generated mice that were heterozygous for the *Cre*-excised (Δ -cat) allele and bearing the *Cre*-transgene (Δ -cat/wt;*Cre*^{+/+}). *Cre* expression in the progeny commences in the blastocyst, so genomic sequences flanked by *loxP* sites are ablated in all cells, including germ cells. Litters containing both control and Δ -cat mice were generated by breeding Δ -cat/wt;*Cre*^{+/+} mice with homozygous *loxP/loxP* mice. In these studies the *loxP/wt*;*Cre*⁻ mice served as controls for the Δ -cat/ Δ -cat;*Cre*⁺ mice.

Mouse genotypes were confirmed by Southern blot analysis of genomic DNA (Fig. 1c). The *Adamts4* complementary DNA probe detected bands corresponding to the *Adamts4* floxed (9.2 kilobases (kb)), wild-type (7.2 kb) and Δ -cat (6.8 kb) alleles. The *Adamts5* cDNA probe detected bands corresponding to the *Adamts5* floxed (11 kb), wild-type (9 kb) and Δ -cat (8.8 kb) alleles. For routine genotyping (Fig. 1d), genomic DNA was extracted from tail biopsies and analysed by polymerase chain reaction (PCR) using primers 1 and 2 for *Adamts4*, and primers 4 and 5 for *Adamts5*. Primers used to detect expression of the *Cre* transgene were included in this PCR analysis. The results show that *Cre* recombinase efficiently ablated the floxed sequence in both genes to create the Δ -cat/ Δ -cat genotype (Δ -cat mice, Fig. 1c, d).

Mouse tissues were analysed by PCR with reverse transcription (RT-PCR) for the presence of *Adamts4* or *Adamts5* messenger RNA transcripts, using primers flanking the ablated sequences. *Adamts4* and 5 were widely expressed (Fig. 1e, f), but the expression of *Adamts4* was low in spine, lung, liver and skin, and absent in the spleen (Fig. 1e). The PCR bands detected in Δ -cat mice were of the predicted size, confirming that *Cre*-mediated excision of the exons encoding the catalytic domain was efficient in both lines of Δ -cat mice. Sequencing of the Δ -cat PCR products confirmed that excision of the floxed exons created an in-frame deletion in both mouse lines. The mutant genes were transcribed, and not removed by nonsense-mediated mRNA decay¹⁵, further confirming that the mutations were in-frame.

Homozygous ADAMTS4 and ADAMTS5 Δ -cat mice were viable, fertile and phenotypically normal; they were indistinguishable from their littermate controls based on gross histological appearance (data not shown).

To determine whether ablation of *Adamts4* or 5 activity caused a compensatory increase in the expression of other aggrecanases, real-time RT-PCR was used to measure the levels of mRNA for *Adamts1*, 4, 5 and 9 extracted from femoral head cartilage. There were no significant changes ($P > 0.05$, Mann-Whitney U-test) in mRNA expression of any ADAMTS enzyme in either the ADAMTS4 or ADAMTS5 Δ -cat mice, compared with littermate controls (Table 1).

A history of *in vivo* aggrecanase activity can be determined by detecting the C-terminal NVTEGE neoepitope (newly created epitope) present in cartilage extracts. This neoepitope is created upon ADAMTS cleavage of the aggrecan core protein at the NVTEGE³⁷³/A³⁷⁴LGSVI bond (where superscript numbers denote the amino acid residue position). Western blot analysis of cartilage extracts showed that the level of NVTEGE neoepitope was reduced in cartilage from both ADAMTS4 and ADAMTS5 Δ -cat mice, compared with control mice (Fig. 1g). This is consistent with

both enzymes having a role in normal aggrecanolytic activity in the mouse. The loss of epitope was more substantial in extracts from ADAMTS5 Δ -cat mice than in ADAMTS4 Δ -cat mice.

Interleukin-1 is a major inflammatory cytokine *in vivo*, and is able to promote aggrecan loss from cartilage explants *in vitro*. Femoral head cartilage was placed in explant culture and stimulated with interleukin-1 α (IL-1 α) to promote aggrecan loss (Fig. 2). As expected, IL-1 α stimulated the release of aggrecan from control cartilage into the conditioned medium, by 2–3-fold (Fig. 2a). IL-1 α

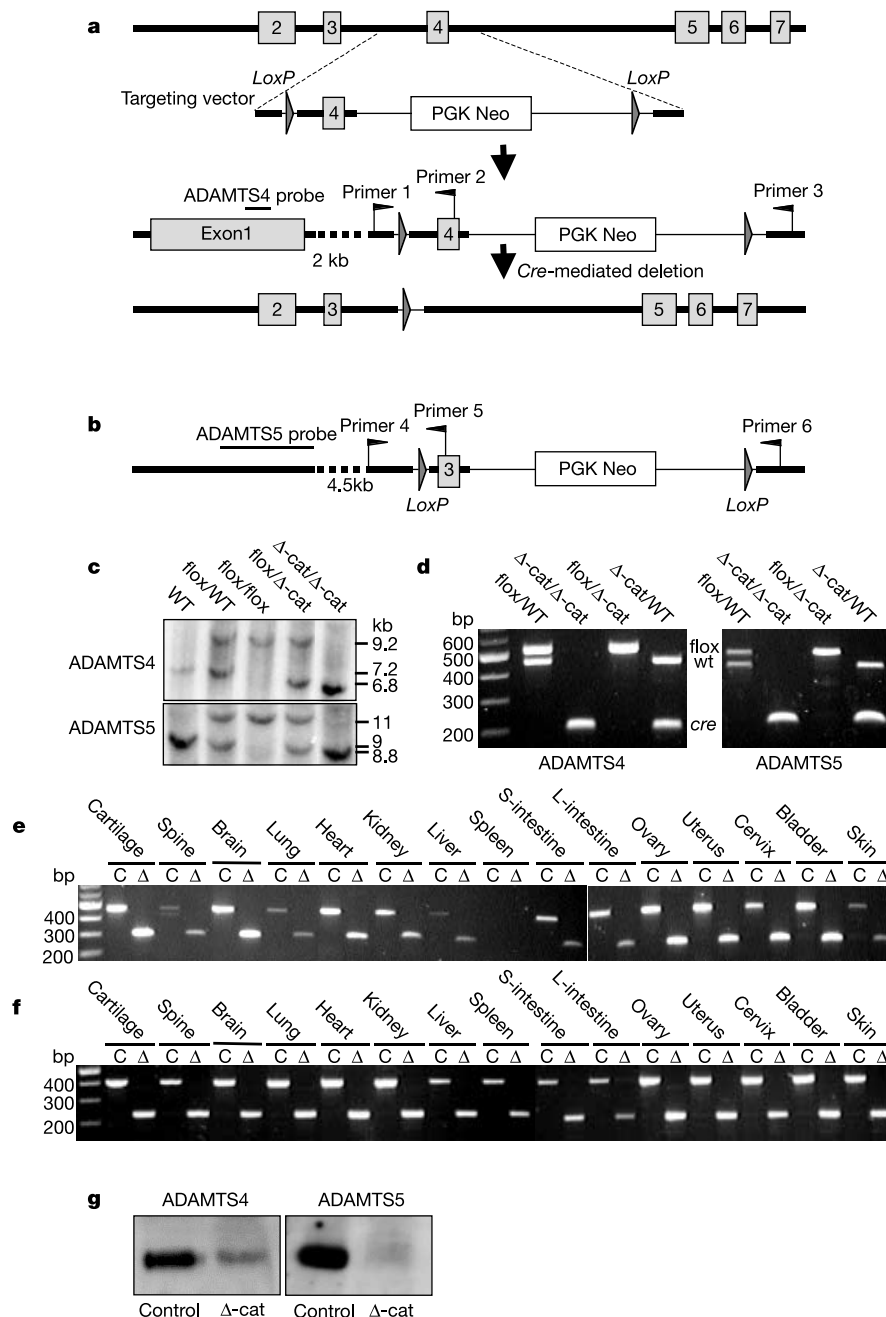


Figure 1 Generation of the ADAMTS4 and ADAMTS5 Δ -cat mice. **a**, Targeting strategy for deletion of exon 4 of the *Adamts4* gene, including positions of genotyping primers and Southern blot probe. PGK, phosphoglycerine kinase promoter used to drive expression of the neomycin-resistance gene (neo). **b**, Representation of the modified *Adamts5* allele. **c**, Southern blot of genomic DNA. WT, wild type. **d**, PCR analysis of genomic DNA from routine genotyping. A second PCR was performed using primers flanking the modified

regions of the two genes (primers 1 and 3 for *Adamts4* and primers 4 and 6 for *Adamts5*) to detect the presence of the Δ -cat allele (results not shown). **e**, **f**, RT-PCR analysis of *Adamts4* and *Adamts5* expression, respectively, in tissues from control (C) and Δ -cat (Δ) mice. **g**, Western blot of the NVTEGE neoepitope in control, ADAMTS4 and ADAMTS5 Δ -cat mice.

Table 1 Quantification of mRNA levels of ADAMTS enzymes in cartilage using real-time RT-PCR

Δ -cat mice	mRNA	ΔC_T (control) $\pm$ s.e.m.	ΔC_T (Δ -cat) $\pm$ s.e.m.	mRNA levels in Δ -cat mice relative to control mice
ADAMTS4	ADAMTS1	16.0 $\pm$ 0.2	15.3 $\pm$ 0.1	0.6
	ADAMTS4	19.8 $\pm$ 0.1	19.5 $\pm$ 0.2	0.8
	ADAMTS5	18.7 $\pm$ 0.0	19.2 $\pm$ 0.1	1.4
	ADAMTS9	15.2 $\pm$ 0.4	15.3 $\pm$ 0.4	1.1
ADAMTS5	ADAMTS1	18.5 $\pm$ 0.0	18.4 $\pm$ 0.0	0.9
	ADAMTS4	19.1 $\pm$ 0.1	19.0 $\pm$ 0.1	0.9
	ADAMTS5	17.0 $\pm$ 0.2	17.0 $\pm$ 0.0	1.0
	ADAMTS9	16.8 $\pm$ 0.1	17.2 $\pm$ 0.1	1.3

Total RNA was extracted and prepared for real-time PCR using Taqman technology. The level of each ADAMTS mRNA was quantified relative to an 18S rRNA internal reference (ΔC_T value). Each value is the mean of three independent pools of femoral head cartilage. The difference between Δ -cat and control mice was calculated using the formula $2^{\Delta C_T(\Delta\text{-cat}) - \Delta C_T(\text{control})}$.

also stimulated release of aggrecan from ADAMTS4 Δ -cat cartilage by 2–3-fold; however, no aggrecan release from ADAMTS5 Δ -cat cartilage was observed upon IL-1 α stimulation, indicating that ADAMTS5 activity accounts almost exclusively for the IL-1 α effect seen in control cartilage.

To examine this further, we performed western blotting using three different antibodies recognizing aggrecanase neopeptides, in order to detect aggrecanase cleavage products present in the conditioned medium or extracts of cultured femoral heads. The mouse NVTEGE and ALGSV neopeptides are created in equimolar amounts by aggrecanase cleavage at the Glu³⁷³/Ala³⁷⁴ bond. The NVTEGE epitope present in tissue extracts (Fig. 2b), and the ALGSV epitope released into the medium (Fig. 2c), were at low levels or absent in unstimulated control cultures (lanes 1, 3, 5, 7), but increased in response to stimulation with IL-1 α in control (lanes 2, 6) and ADAMTS4 Δ -cat (lane 4) cartilage. In contrast, IL-1 α was unable to stimulate production of either the NVTEGE or the ALGSV neopeptides in cartilage from the ADAMTS5 Δ -cat mice (lane 8, Fig. 2b, c). The AGEV epitope is a C-terminal neopeptide generated by ADAMTS cleavage at the Glu¹⁵⁷²/Ala¹⁵⁷³ bond of aggrecan. AGEV immunoreactivity was present in medium from unstimulated control (lanes 1 and 5, Fig. 2d) and ADAMTS4 Δ -cat (lane 3, Fig. 2d) cartilage, but was absent in medium from ADAMTS5 Δ -cat cartilage (lane 7, Fig. 2d). IL-1 α stimulation generated a small increase in the AGEV epitope in both control and Δ -cat mice. Given that only the AGEV neopeptide was increased by IL-1 α treatment in ADAMTS5 Δ -cat cartilage, there may be a residual, IL-1 α -inducible aggrecanase activity in ADAMTS5 Δ -cat cartilage that cleaves at the Glu¹⁵⁷²/Ala¹⁵⁷³ bond, but not at the major Glu³⁷³/Ala³⁷⁴ bond; this activity could be due to ADAMTS4. The NVTEGE, ALGSV and AGEV western blot results are consistent with the results for total aggrecan release (Fig. 2a), and strongly suggest that in the mouse, IL-1 α acts to increase ADAMTS5 but not ADAMTS4 activity. Real-time RT-PCR analysis of cartilage from control mice revealed that IL-1 α increased ADAMTS5 mRNA expression by 18-fold, and increased ADAMTS4 mRNA expression threefold. Very similar results were obtained in cartilage from both ADAMTS4 and ADAMTS5 Δ -cat mice, demonstrating that IL-1 α strongly stimulates ADAMTS5 expression at the transcriptional level.

Other investigators have reported that ADAMTS4 is not responsible for cartilage damage in a mouse model of osteoarthritis¹⁶. This is consistent with the results above (Fig. 2), which show that ADAMTS4 is not the major aggrecanase in mouse articular cartilage *in vitro*. We used a model of inflammatory arthritis to determine whether ablation of ADAMTS5 activity could protect against aggrecan loss and cartilage damage in the mouse. Methylated bovine serum albumin (mBSA) was injected into the knee joints of primed mice in order to induce monoarticular arthritis. There were no significant differences between control and ADAMTS5 Δ -cat mice in the severity of synovitis, joint space exudate, pannus-mediated erosion or bone destruction ($P > 0.05$). The total histological scores (mean $\pm$ s.e.m.) were 7.9 ± 0.3 for control mice and 7.5 ± 0.3 for ADAMTS5 Δ -cat mice, out of a maximum possible

score of 12. Joints receiving saline were used as negative controls and showed no signs of inflammation or cartilage damage (Fig. 3b).

Aggrecan loss from cartilage was assessed by loss of toluidine blue staining in sections of the tibiofemoral joint. Aggrecan loss in the

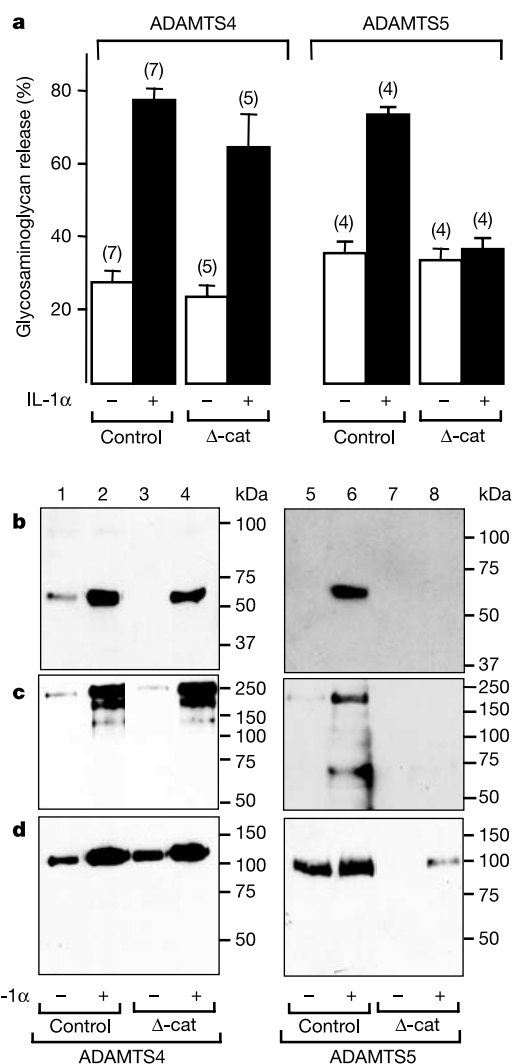


Figure 2 IL-1 α stimulation of aggrecan release in cartilage explant cultures. **a**, Aggrecan released into the culture medium was calculated as the percentage glycosaminoglycan in medium (glycosaminoglycan in the medium divided by total glycosaminoglycan present in the medium + extracts + papain digests $\times$ 100). Results show means $\pm$ s.e.m. of two experiments, with the total number of replicates given in parentheses. **b**, Aggrecan fragments present in cartilage extracts were detected by western blotting with an antibody recognizing the aggrecanase-generated neopeptide NVTEGE. **c**, **d**, Aggrecan fragments released into the medium were detected by western blotting using antibodies specific for the aggrecanase-generated neopeptides ALGSV (**c**) and AGEV (**d**).

ADAMTS5 Δ -cat mice was significantly less (Fig. 3a, e) than in the controls ($P < 0.01$, Fig. 3a, d). This effect is most clearly seen here in the tibial cartilage. The score (mean $\pm$ s.e.m.) for the control joints ($n = 14$) was 1.94 ± 0.13 , whereas the score for the ADAMTS5- Δ -cat joints ($n = 13$) was 1.15 ± 0.15 (Fig. 3a). Tissue sections were also monitored for cartilage erosion and evidence of surface roughening and fibrillation. Five of fourteen (36%) joints from control mice were positive for cartilage erosion (Fig. 3f), whereas only one of thirteen (7%) joints from the ADAMTS5 Δ -cat joints showed cartilage erosion. A typical ADAMTS5 Δ -cat joint section is shown in Fig. 3g.

Our results indicate that ablation of ADAMTS5 activity in the mouse is sufficient to protect against aggrecan loss and cartilage erosion in a model of inflammatory arthritis. The results are consistent with other mouse studies showing that 'aggrecanase activity' is involved in initiating early aggrecan loss and cartilage damage^{17–19}, but that ADAMTS4 is not the enzyme involved¹⁶. In humans, ADAMTS5 is expressed at higher levels than ADAMTS4

(refs 12, 20) in both normal and arthritic cartilage. However, the change in expression with disease is less clear because both increases¹² and decreases²⁰ in ADAMTS4 and 5 have been reported for osteoarthritic cartilage, compared with normal cartilage. These differences may be related to the stage of disease.

This is the first report of a single enzyme deficiency that diminishes cartilage destruction in the mouse. Further studies are required to determine whether ADAMTS5 is also important in human diseases. Our data suggest that ADAMTS5 may be a suitable target for the development of new drugs designed to inhibit cartilage destruction in arthritis. □

Methods

Genotyping

Genotyping was done by PCR using genomic DNA extracted from tail biopsies by proteinase K digestion. The primers used to amplify the *Cre* transgene were 5'-CGCTAAGGATGAC TCTGGT-3' (forward) and 5'-CTAATCGCCATCTCCAGCAGG-3' (reverse). The sequences of the primers shown in Fig. 1a, b are: Primer 1, 5'-CCCGGCTCATGGATTC TTC-3'; Primer 2, 5'-GCTCTTCAGGGTCCACATG-3'; Primer 3, 5'-CCGTTTCAATCT GGGACAGT-3'; Primer 4, 5'-TGGTGAGACTATGACATAATCC-3'; Primer 5, 5'-GCTTGCTTCTGTAGTACCG-3'; Primer 6, 5'-TGGTCTCCCACT CTTTCC-3'.

Southern blotting

Southern blot analysis was done using genomic DNA extracted from tail biopsies. DNA was digested with the restriction enzyme *Bam*HI to examine the *Adams4* gene, or with *Bgl*II to examine the *Adams5* gene, then electrophoresed on 0.3% agarose gels for 48 h and transferred to PVDF membranes. The membranes were screened with ³²P-dCTP cDNA probes and exposed to X-ray film.

RT-PCR

RT-PCR of *Adams4* and 5 mRNA expression was done with total RNA extracted from tissues, using primers flanking the ablated exons. The primers for *Adams4* were 5'-ATGTGGGCACAGTGTGTGAT-3' (forward) and 5'-CAAGGTGAGTGCTTCGTCTG-3' (reverse), and for *Adams5* the primers were 5'-GGCATCATTCATGTGACACC-3' (forward) and 5'-CGAGTACTCAGGCCCAATG-3' (reverse).

In vitro analysis of aggrecan catabolism

In vitro aggrecan catabolism was analysed by culturing femoral head cartilage with and without IL-1 α . Intact femoral heads were harvested from the hip joints of 3–4-week-old mice by shearing the capital femoral physis from the underlying metaphyseal bone, using pressure with forceps⁶. The explants were cultured at 37 °C with 5% CO₂ in DMEM medium containing 10% FCS, antibiotics, 2 mM L-glutamine and 20 mM HEPES. After 2 days the explants were stimulated for 3 days with 10 ng ml⁻¹ human IL-1 α in serum-free medium. Control cartilage was cultured in serum-free medium without IL-1 α . At the end of the culture period the cartilage was blotted dry, weighed, and aggrecan and G1-enriched aggrecan fragments were extracted using 4 M guanidine hydrochloride buffer (pH 5.8), containing proteinase inhibitors. The extract was dialysed and the concentration of sulphated glycosaminoglycan determined using the 1,9-dimethylmethylene blue assay²¹. Samples were deglycosylated using chondroitinase ABC in the presence of proteinase inhibitors, before SDS-PAGE and western blotting. The papain-digested tissue residue and the conditioned medium were assayed for sulphated glycosaminoglycans.

Real-time Taqman PCR

Real-time Taqman PCR was used to detect the expression of mRNAs for *Adams1*, *Adams4*, *Adams5* and *Adams9*. Primer and probe sets were designed and synthesized using the Applied Biosystems 'Primer-by-Design' service. Each probe spanned an exon-exon boundary. The sequences of each primer/probe set were: *Adams1* forward primer, 5'-CCTGTGAAGCCAAAGGCATTG-3' (exon 7); *Adams1* reverse primer, 5'-TGCACACAGACAGAGGTAGAGT-3' (exon 8); *Adams1* Taqman probe, 5'-CAGCCC AAGGTTGTAGATG-3'; *Adams4* forward primer, 5'-CAAGCAGTCGGGCTCCTT-3' (exon 8); *Adams4* reverse primer, 5'-GATCGTGACCACATCGCTGTA-3' (exon 9); *Adams4* Taqman probe, 5'-TCCATACCTGAATTTTGTG-3'; *Adams5* forward primer, 5'-CCAGTTGTACAAAGATTATCGGAACCT-3' (exon 7); *Adams5* reverse primer, 5'-GTTGCTCCTCAGGGATCCT-3' (exon 8); *Adams5* Taqman probe, 5'-TCAGTATA ACCCTTGCTTTTTT-3'; *Adams9* forward primer, 5'-AGCAGCTGCGGTCATCAG-3' (exon 2); *Adams9* reverse primer, 5'-ACAGACTGCAGTGGTTCAATGAAATA-3' (exon 3); *Adams9* Taqman probe, 5'-AACGTGCCATCATTC-3'. The primer/probe set for amplification of 18S ribosomal RNA was purchased from Applied Biosystems. Multiplex PCR was done in an ABI Prism 7700 Sequence Detection System, using primer/probe sets to the target (ADAMTS) and reference (18S rRNA) in the same reaction. The relative amount of mRNA expression between control and Δ -cat mice was determined by the comparative cycle threshold (C_T) method²², using the formula $2^{(\Delta C_T(\Delta\text{-cat}) - \Delta C_T(\text{control}))}$, where ΔC_T is calculated by subtracting the C_T value of the reference RNA from the C_T value of the target RNA.

Antigen-induced arthritis

A modification of the antigen-induced arthritis model^{23,24} was used to assess the *in vivo* consequence of ablating ADAMTS5 activity. On day 0, 8-week-old mice were primed with

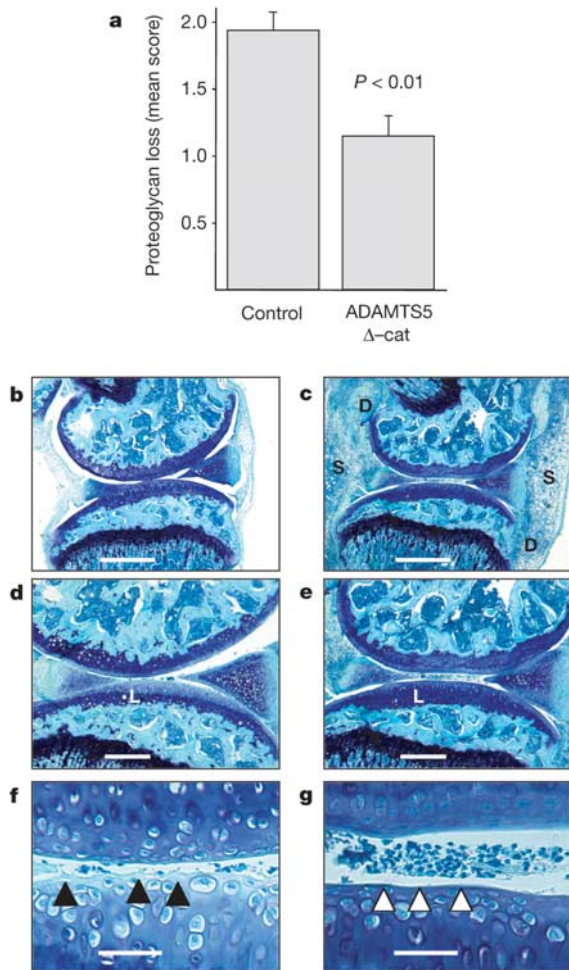


Figure 3 Antigen-induced arthritis in control and ADAMTS5-deficient mice. **a**, Sections of the tibiofemoral joints from control ($n = 14$), and ADAMTS5 Δ -cat ($n = 13$) mice were scored for aggrecan loss, and the data analysed using the Mann-Whitney U-test. Error bars indicate s.e.m. **b, c**, Toluidine blue-stained sections from ADAMTS5 Δ -cat mice injected intra-articularly with saline (**b**) or mBSA (**c**) showing areas of synovial hyperplasia (S) and bone and marginal zone destruction (D) ($\times 40$ original magnification). **d, e**, Tibiofemoral joints ($\times 100$ original magnification) injected with mBSA, showing regions of aggrecan loss (L) in control mice (**d**) compared with the same region in ADAMTS5 Δ -cat mice (**e**). **f, g**, Tibiofemoral joints ($\times 400$ original magnification) injected with mBSA, showing cartilage erosion (arrowheads) in control mice (**f**), compared with the same region in ADAMTS5 Δ -cat mice (**g**). Scale bar, 500 μ m (**b, c**), 250 μ m (**d, e**) or 50 μ m (**f, g**).

an intradermal injection of 50 µg mBSA in Freund's Complete Adjuvant at two sites near the base of the tail. Arthritis was induced on day 7, by intra-articular injection of 10 µl of 20 mg ml⁻¹ mBSA directly into both knee joints. As a negative control, some mice received 10 µl mBSA in the right knee only, and 10 µl saline in the left knee; inflammation ensued in the mBSA-injected knee only. The mice were killed on day 10 (3 days after induction of arthritis, when aggrecan loss is maximal¹⁸) and the knee joints were processed for histology. Toluidine blue-stained slides, containing 3–5 consecutive sections from a single joint, were blinded and scored independently by five examiners. The slides were scored for aggrecan loss using a four-point scale, where 0 = no loss of staining intensity in non-calcified articular cartilage (NCAC) and calcified articular cartilage, 1 = localized decrease, but not total loss of staining in NCAC, 2 = widespread decrease in staining and localized complete loss of staining in NCAC, and 3 = widespread complete loss of staining in NCAC. The femur and the tibia were evaluated in each section. The sections were also assessed for evidence of cartilage fibrillation. The inflammatory and destructive parameters of synovitis, joint space exudates, and pannus-mediated erosion and bone destruction were scored on a scale from 0–3, with 3 being the maximum score. Total histological scores were analysed for significance using Mann-Whitney U-tests.

Received 6 January; accepted 2 February 2005; doi:10.1038/nature03417.

- Kuno, K. *et al.* Molecular cloning of a gene encoding a new type of metalloproteinase-disintegrin family protein with thrombospondin motifs as an inflammation associated gene. *J. Biol. Chem.* **272**, 556–562 (1997).
- Kuno, K. *et al.* ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan. *FEBS Lett.* **478**, 241–245 (2000).
- Rodriguez-Manzanique, J. C. *et al.* ADAMTS1 cleaves aggrecan at multiple sites and is differentially inhibited by metalloproteinase inhibitors. *Biochem. Biophys. Res. Commun.* **293**, 501–508 (2002).
- Somerville, R. P. *et al.* Characterization of ADAMTS-9 and ADAMTS-20 as a distinct ADAMTS subfamily related to *Caenorhabditis elegans* GON-1. *J. Biol. Chem.* **278**, 9503–9513 (2003).
- Collins-Racie, L. A. *et al.* ADAMTS-8 exhibits aggrecanase activity and is expressed in human articular cartilage. *Matrix Biol.* **23**, 219–230 (2004).
- Little, C. B. *et al.* ADAMTS-1 knockout mice do not exhibit abnormalities in aggrecan turnover *in vitro* or *in vivo*. *Arthritis Rheum.* (in the press).
- Abbaszade, I. *et al.* Cloning and characterization of ADAMTS11, an aggrecanase from the ADAMTS family. *J. Biol. Chem.* **274**, 23443–23450 (1999).
- Tortorella, M. D., Malfait, A. M., Deccico, C. & Arner, E. The role of ADAM-TS4 (aggrecanase-1) and ADAM-TS5 (aggrecanase-2) in a model of cartilage degradation. *Osteoarthritis Cartilage* **9**, 539–552 (2001).
- Vankemmelbeke, M. N. *et al.* Expression and activity of ADAMTS-5 in synovium. *Eur. J. Biochem.* **268**, 1259–1268 (2001).
- Yamanishi, Y. *et al.* Expression and regulation of aggrecanase in arthritis: the role of TGF-β. *J. Immunol.* **168**, 1405–1412 (2002).
- Ilic, M. Z. *et al.* Bovine joint capsule and fibroblasts derived from joint capsule express aggrecanase activity. *Matrix Biol.* **19**, 257–265 (2000).
- Bau, B. *et al.* Relative messenger RNA expression profiling of collagenases and aggrecanases in human articular chondrocytes *in vivo* and *in vitro*. *Arthritis Rheum.* **46**, 2648–2657 (2002).
- Koshy, P. J. *et al.* The modulation of matrix metalloproteinase and ADAM gene expression in human chondrocytes by interleukin-1 and oncostatin M: A time-course study using real-time quantitative reverse transcription-polymerase chain reaction. *Arthritis Rheum.* **46**, 961–967 (2002).
- Schwenk, F., Baron, U. & Rajewsky, K. A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23**, 5080–5081 (1995).
- Mendell, J. T., Sharifi, N. A., Meyers, J. L., Martinez-Murillo, F. & Dietz, H. C. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nature Genet.* **36**, 1073–1078 (2004).
- Glasson, S. S. *et al.* Characterization of and osteoarthritis susceptibility in ADAMTS-4-knockout mice. *Arthritis Rheum.* **50**, 2547–2558 (2004).
- van Meurs, J. B. *et al.* Increased vulnerability of postarthritic cartilage to a second arthritic insult: accelerated MMP activity in a flare up of arthritis. *Ann. Rheum. Dis.* **58**, 350–356 (1999).
- van Meurs, J. B. *et al.* Kinetics of aggrecanase- and metalloproteinase-induced neopeptides in various stages of cartilage destruction in murine arthritis. *Arthritis Rheum.* **42**, 1128–1139 (1999).
- van Meurs, J. B. *et al.* Interleukin-1 receptor antagonist prevents expression of the metalloproteinase-generated neopeptide VDIPEN in antigen-induced arthritis. *Arthritis Rheum.* **41**, 647–656 (1998).
- Kevorkian, L. *et al.* Expression profiling of metalloproteinases and their inhibitors in cartilage. *Arthritis Rheum.* **50**, 131–141 (2004).
- Farndale, R. W., Sayers, C. A. & Barrett, A. J. A direct spectrophotometric microassay for sulfated glycosaminoglycans in cartilage cultures. *Connect. Tissue Res.* **9**, 247–248 (1982).
- Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).
- Mould, A. W. *et al.* Vegf gene knockout mice display reduced pathology and synovial angiogenesis in both antigen-induced and collagen-induced models of arthritis. *Arthritis Rheum.* **48**, 2660–2669 (2003).
- Glant, T. T. *et al.* Cartilage-specific constitutive expression of TSG-6 protein (product of tumor necrosis factor α-stimulated gene 6) provides a chondroprotective, but not antiinflammatory, effect in antigen-induced arthritis. *Arthritis Rheum.* **46**, 2207–2218 (2002).

Acknowledgements The authors are grateful to Johnson & Johnson Pharmaceutical Research and Development (San Diego) for donating the floxed mice. We thank E. Arner and M. Tortorella for providing the anti-AGEC antibody. We also gratefully acknowledge the mouse husbandry support provided by the Disease Models Unit of the Murdoch Childrens Research Institute. The work was funded by the National Health and Medical Research Council (Australia), and the Murdoch Childrens Research Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.J.F. (amanda.fosang@mcri.edu.au).

Cyclophilin D-dependent mitochondrial permeability transition regulates some necrotic but not apoptotic cell death

Takashi Nakagawa^{1,2,4}, Shigeomi Shimizu^{1,4}, Tetsuya Watanabe³, Osamu Yamaguchi³, Kinya Otsu³, Hirohisa Yamagata¹, Hidenori Inohara², Takeshi Kubo² & Yoshihide Tsujimoto^{1,4}

¹Laboratory of Molecular Genetics, Department of Post-Genomics and Diseases,

²Department of Otolaryngology and Sensory Organ Surgery and ³Department of Internal Medicine and Therapeutics, Osaka University Medical school, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

⁴Solution-Oriented Research for Science and Technology (SORST), Japan Science and Technology Corporation, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

Mitochondria play an important role in energy production, Ca²⁺ homeostasis and cell death. In recent years, the role of the mitochondria in apoptotic and necrotic cell death has attracted much attention^{1,2}. In apoptosis and necrosis, the mitochondrial permeability transition (mPT), which leads to disruption of the mitochondrial membranes and mitochondrial dysfunction, is considered to be one of the key events, although its exact role in cell death remains elusive. We therefore created mice lacking cyclophilin D (CypD), a protein considered to be involved in the mPT, to analyse its role in cell death. CypD-deficient mice were developmentally normal and showed no apparent anomalies, but CypD-deficient mitochondria did not undergo the cyclosporin A-sensitive mPT. CypD-deficient cells died normally in response to various apoptotic stimuli, but showed resistance to necrotic cell death induced by reactive oxygen species and Ca²⁺ overload. In addition, CypD-deficient mice showed a high level of resistance to ischaemia/reperfusion-induced cardiac injury. Our results indicate that the CypD-dependent mPT regulates some forms of necrotic death, but not apoptotic death.

The mitochondrial permeability transition (mPT) is a regulated Ca²⁺-dependent increase in the permeability of the mitochondrial membrane, which results in a loss in mitochondrial membrane potential (ΔΨ), mitochondrial swelling and rupture of the outer membrane³. The mPT is thought to occur after the opening of a channel, termed the permeability transition pore, which putatively consists of a voltage-dependent anion channel, an adenine nucleotide translocator, CypD, and some other molecule(s)⁴; however, an essential role for the adenine nucleotide translocator in the mPT is a matter of recent controversy^{5,6}.

CypD is a mitochondrial member of the cyclophilin family of peptidyl prolyl-*cis*, *trans*-isomerases (PPIases) and has a crucial role in protein folding⁷. It has been suggested that CypD is involved in regulating the mPT, on the basis of the observation that cyclosporin A (CsA), a specific inhibitor of cyclophilin family activity, blocks the mPT⁸. A CsA-insensitive mPT has also been suggested, although the molecular mechanism is completely unknown⁹. It has been shown that some forms of apoptosis are significantly inhibited by CsA, suggesting a role for CsA-sensitive mPT in apoptosis^{2,4}. The mPT is also implicated in the remodelling of mitochondrial structure with mobilization of cytochrome *c* stores in cristae during apoptosis¹⁰. To determine whether CypD has a crucial role in the CsA-sensitive mPT, and to investigate whether the mPT is a key regulator of cell death, we created CypD-deficient mice by gene targeting (see Supplementary Fig. 1a–c). The absence of cyclophilin D protein in CypD-deficient mice was verified by western blotting (see Fig. 1a and Supplementary Fig. 1d). CypD-deficient mice were born at the

expected mendelian frequency, developed normally, and did not have any detectable phenotypic anomalies.

To examine the role of CypD in the mPT, mitochondria were isolated from the livers of CypD-deficient mice and control littermates. The mitochondria showed no significant change in respiration rate in the absence of CypD (see Supplementary Fig. 2). As shown in Fig. 1a, PPIase activity was absent in CypD-deficient mitochondria, but not in control mitochondria, indicating that CypD is the major PPIase in the mitochondria. When control

mitochondria were treated with $50 \mu\text{M}$ Ca^{2+} , the mPT occurred, as shown by mitochondrial swelling (Fig. 1b) and loss of $\Delta\Psi$ (Fig. 1c). These phenomena were not observed in CypD-deficient mitochondria (Fig. 1b, c). To examine the extent of mPT inhibition as a consequence of CypD-deficiency, successive doses of Ca^{2+} were added to the mitochondria. $\Delta\Psi$ was lost after two additions of $50 \mu\text{M}$ Ca^{2+} to control mitochondria, but it was still maintained after seven additions of Ca^{2+} to CypD-deficient mitochondria (Fig. 1d). We investigated whether the absence of Ca^{2+} -induced

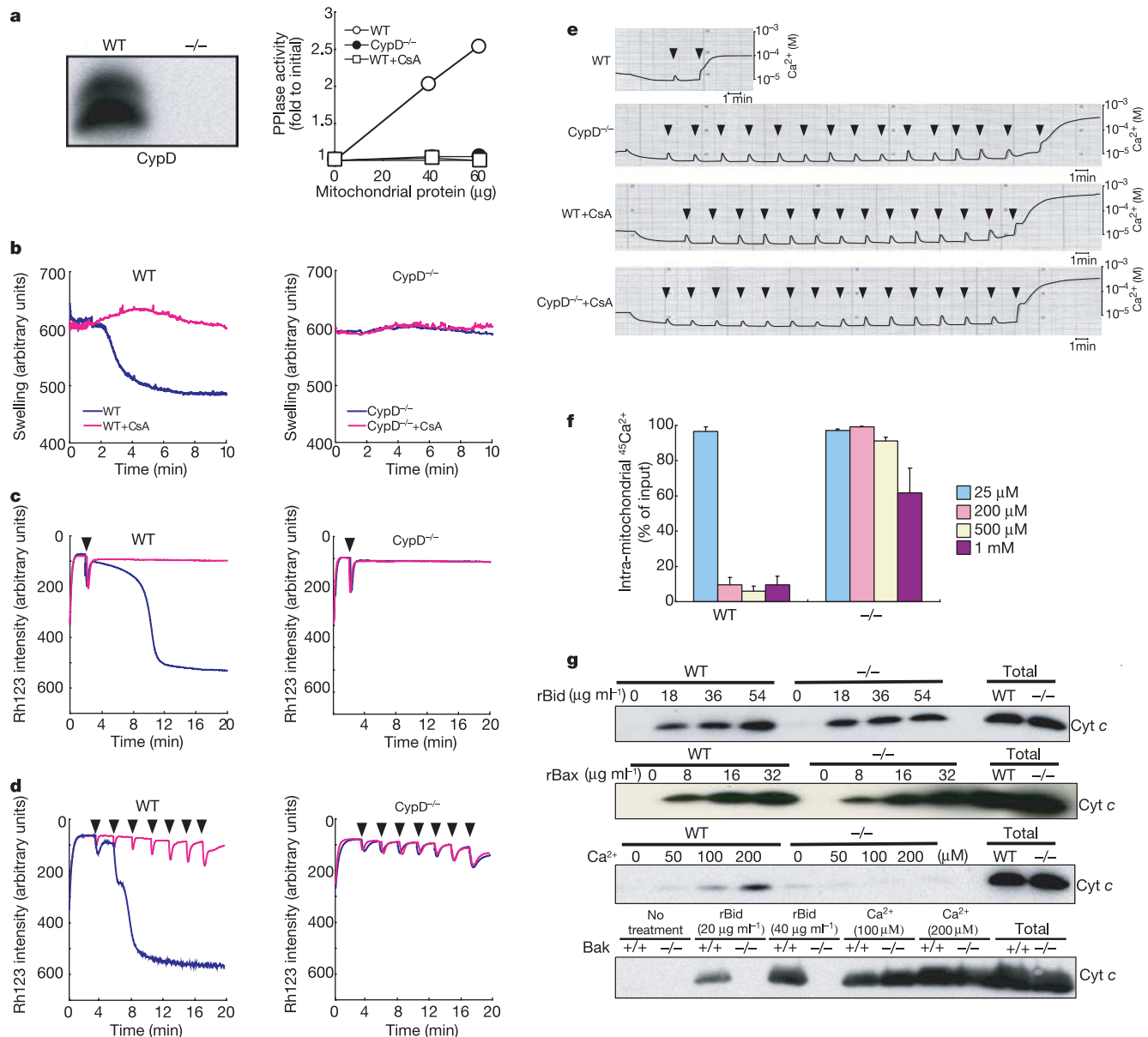


Figure 1 Absence of mPT in CypD-deficient (CypD^{-/-}) mitochondria. **a**, Absence of CypD protein and PPIase activity in CypD^{-/-} mitochondria. WT, wild type. **b**, **c**, Absence of mPT in CypD^{-/-} mitochondria. Isolated wild-type (WT, left column) or CypD^{-/-} (right column) mitochondria were incubated with $50 \mu\text{M}$ Ca^{2+} in the presence (pink) or absence (blue) of $1 \mu\text{M}$ CsA, and monitored for **(b)** swelling (by light scatter) or **(c)** $\Delta\Psi$ (by Rh123 intensity), see Methods. Loss of $\Delta\Psi$ causes release of Rh123 from the mitochondria, resulting in increased Rh123 intensity. **d**, **e**, CypD deficiency prevents Ca^{2+} -induced $\Delta\Psi$ loss without altering Ca^{2+} uptake. Isolated mitochondria were successively treated with $50 \mu\text{M}$ Ca^{2+} (indicated by arrowheads) in the presence (pink) or absence (blue) of $1 \mu\text{M}$ CsA, and $\Delta\Psi$ (**d**) and extra-mitochondrial Ca^{2+} (**e**) were monitored.

f, Accumulation of Ca^{2+} in the mitochondria as a result of CypD deficiency. Mitochondria were incubated with the indicated concentrations of $^{45}\text{Ca}^{2+}$ for 25 min and intra-mitochondrial Ca^{2+} was measured. Data shown as mean $\pm$ s.e.m. **g**, Lack of Ca^{2+} -induced, but not rBid- and rBax-induced, cytochrome *c* release in CypD^{-/-} mitochondria. WT, Bak^{-/-} and CypD^{-/-} mitochondria were incubated with Ca^{2+} , rBid, and rBax at the indicated concentrations. After 30 min, samples were centrifuged and aliquots of supernatants were subjected to western blot analysis for cytochrome *c*. 'Total' represents the total amount of cytochrome *c* found in an equivalent aliquot of the mitochondria.

mPT in CypD-deficient mitochondria was due to disturbance of Ca^{2+} uptake. As shown in Fig. 1e, the extra-mitochondrial Ca^{2+} concentration increased transiently, and rapidly returned to basal levels after each successive addition of Ca^{2+} to CypD-deficient mitochondria, indicating normal Ca^{2+} uptake by the mitochondria. Up to concentrations of 500 μM Ca^{2+} , most of the Ca^{2+} was taken up by CypD-deficient mitochondria (Fig. 1e, f). Next, we analysed

CypD-deficient mitochondria in the presence of much higher concentrations of Ca^{2+} , which can induce CsA-insensitive mPT⁹. Addition of more than 1 mM Ca^{2+} induced swelling, collapse of $\Delta\Psi$, and impaired Ca^{2+} uptake even in CypD-deficient mitochondria (Fig. 1f and Supplementary Fig. 3), and all of these events were insensitive to CsA, even when it was added to CypD-deficient mitochondria (see Supplementary Fig. 3). Taken together, these results suggest that the mPT induced by low doses of Ca^{2+} is completely inhibited in CypD-deficient mitochondria, and that CypD is not involved in the CsA-insensitive increase in membrane permeability induced by high doses of Ca^{2+} . Furthermore, the addition of other mPT inducers, such as H_2O_2 and atractyloside, did not trigger the mPT in CypD-deficient mitochondria (see Supplementary Fig. 4).

In many forms of apoptosis, BH3-domain-containing proteins of the Bcl-2 family ('BH3-only' proteins) transduce apoptotic signals to the mitochondria, and induce cytochrome *c* release in a Bax/Bak-dependent manner¹¹. Whether the CsA-sensitive mPT is involved in apoptotic cytochrome *c* release is controversial^{12–15}. To address this question, we added recombinant Bid (rBid), one of the BH3-only proteins, to CypD-deficient and control mitochondria. As shown in Fig. 1g (top panel), CypD deficiency had no effect on rBid-induced cytochrome *c* release, which was different from the result in Bak-deficient mitochondria (Fig. 1g, bottom panel). Bad (another BH3-only protein, data not shown) and rBax (Fig. 1g; second panel), were also found to induce cytochrome *c* release equally in both CypD-deficient and control mitochondria. In contrast, Ca^{2+} -induced cytochrome *c* release was markedly reduced in CypD-deficient mitochondria (Fig. 1g; third panel). Interestingly, Bak deficiency did not have any effect on Ca^{2+} -induced cytochrome *c* release (Fig. 1g, bottom panel). Together, these data indicate that the mPT is involved in cytochrome *c* release induced by Ca^{2+} , but not by pro-apoptotic BH3-only proteins or Bax.

The results described above raise the possibility that the mPT might be involved in cell death due to mPT inducers (including Ca^{2+} overload and reactive oxygen species), but is not involved in the common apoptotic pathway to which BH3-only proteins and Bax/Bak are committed. We first examined the responses of CypD-deficient cells to these death stimuli. As shown in Fig. 2a, control and CypD-deficient thymocytes underwent comparable levels of apoptosis when exposed to various apoptotic stimuli. Similar findings were also obtained when murine embryonic fibroblast cells (MEFs) and hepatocytes from CypD-deficient mice were treated with various apoptotic reagents (Fig. 2b, c), and when these cells were transfected with DNA encoding Bax or a truncated form of Bid (tBid; Fig. 2d, e). Moreover, apoptotic death of intestinal epithelial cells also occurred equally in control and CypD-deficient mice subjected to X-ray irradiation (Fig. 2f). These results indicate that CypD-dependent (CsA-sensitive) mPT is not involved in the common apoptotic pathway.

Next, we tested control and CypD-deficient MEFs for cell death after exposure to H_2O_2 . CypD-deficient MEFs were more resistant to H_2O_2 -induced cell death than control cells as assessed by a CTB (cell titre blue) assay, which measures the metabolic activity of viable cells (Fig. 3a), and by Annexin-V staining (data not shown). H_2O_2 -induced cell death is predominantly due to necrosis, based on the following observations: a lack of caspase activation (Fig. 3b), no effect of a caspase inhibitor (Fig. 3a), early disruption of the plasma membrane (Fig. 3c), and finally, no nuclear or oligonucleosomal DNA fragmentation (Fig. 3c and data not shown). Similar results for H_2O_2 -induced cell death were also obtained using CypD-deficient hepatocytes (Fig. 3d and data not shown).

We also examined the effect of CypD on cell death induced by A23187, a Ca^{2+} ionophore. CypD-deficient hepatocytes showed significant resistance to A23187-induced cell death (Fig. 3e). Like H_2O_2 -induced cell death, A23187-induced cell death was not accompanied by caspase activation (Fig. 3f), so this type of

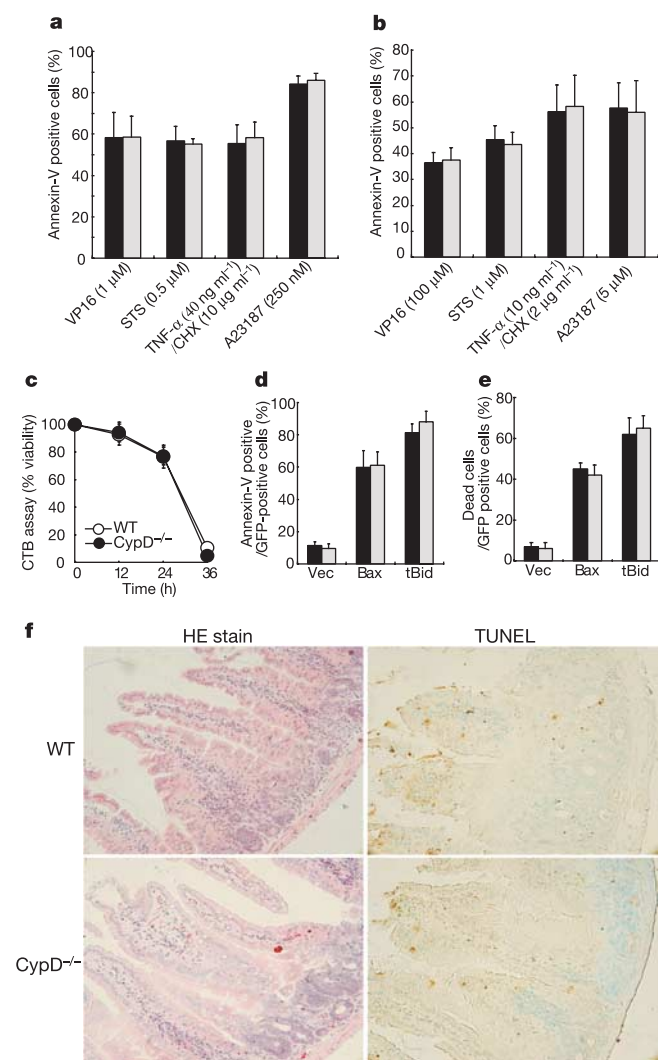


Figure 2 No resistance of CypD^{-/-} cells to multiple apoptotic stimuli. **a, b**, Susceptibility of primary thymocytes and MEFs to various apoptotic stimuli. Wild-type (black) and CypD^{-/-} (grey) thymocytes (**a**) and MEFs (**b**) were exposed to apoptosis-inducing reagents for 24 h, and apoptotic cells were assessed by Annexin-V staining. Reagents: etoposide (VP16), staurosporin (STS), tumour-necrosis factor- α (TNF- α) + cycloheximide (CHX), A23187. **c**, Susceptibility of primary hepatocytes to 10 μM STS. Wild-type and CypD^{-/-} hepatocytes were treated with STS for 36 h and cell death was assessed using the CTB assay. Data shown as mean $\pm$ s.e.m. **d, e**, No effect of CypD deficiency on Bax- or tBid-induced death of MEFs and hepatocytes. Immortalized MEFs (**d**) and primary cultured hepatocytes (**e**) were transiently transfected with DNA for Bax (1 μg) or tBid (1 μg) plus enhanced green fluorescent protein (EGFP, 0.5 μg) for 24 h, and cells were stained with Cy3-conjugated Annexin-V. The percentage of Annexin-V positive cells was calculated relative to all GFP-positive cells. Wild type, black; CypD^{-/-}, grey. Data shown as mean $\pm$ s.e.m. **f**, No effect of CypD deficiency on X-ray-induced apoptosis in the small intestine. Wild-type and CypD^{-/-} mice were exposed to 10 Gy irradiation. After 72 h, a segment of the small intestine was excised and subjected to haematoxylin-eosin (HE) and TUNEL staining.

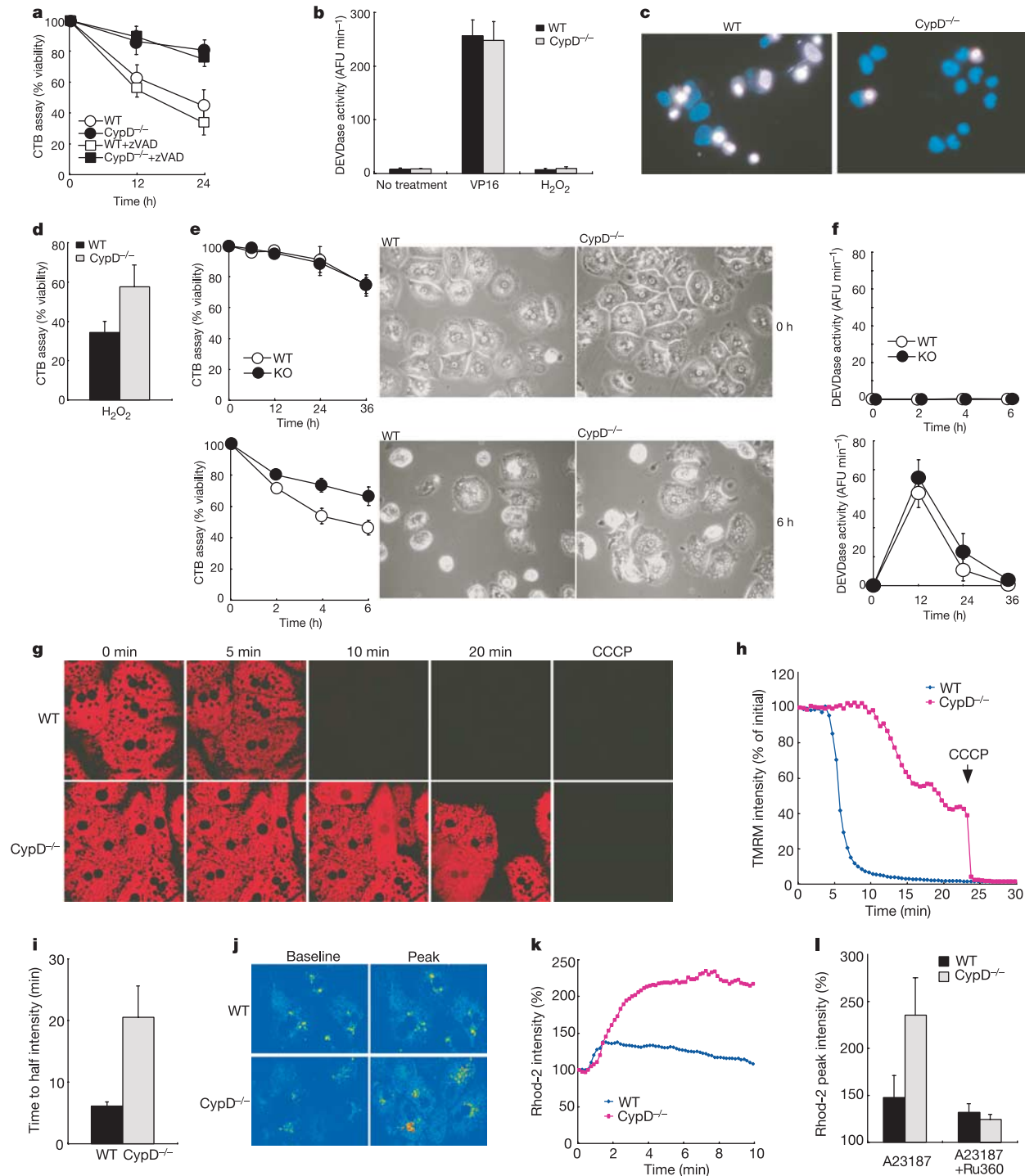


Figure 3 Resistance of CypD^{-/-} cells to necrosis induced by reactive oxygen species and Ca²⁺ overload. **a–d**, Reduction of H₂O₂-induced necrotic cell death by CypD deficiency. Wild-type (WT) and CypD^{-/-} MEFs (**a–c**) and hepatocytes (**d**) were exposed to H₂O₂ (**a–c**, 0.75 mM for 24 h; **d**, 0.5 mM for 4 h), and the extent of cell death was assessed by CTB assay (**a**, **d**). zVAD is a pan-caspase inhibitor. **b**, Caspase activation in wild-type (black) and CypD^{-/-} (grey) hepatocytes after treatment with 0.75 mM H₂O₂ was assessed at 16 h. Treatment with VP16 (100 μM) for 16 h was used as a positive control for caspase activation. Data shown as mean ± s.e.m. **c**, Representative nuclear changes visualized by staining with Hoechst 33342 (blue) and PI (red) at 12 h. **e**, WT (open circles) and CypD^{-/-} (filled circles) hepatocytes left untreated (top panel) or treated with 2 μM A23187 (bottom panel), measured using the CTB assay. Data shown as mean ± s.e.m. Phase contrast microscopy images at 0 h and 6 h are shown. **f**, Assessment of caspase activation in response to 2 μM A23187 (upper panel); treatment with 40 ng ml⁻¹

TNF-α + 10 μg ml⁻¹ CHX was used as a positive control for caspase activation (lower panel). Data shown as mean ± s.e.m. **g–i**, Reduced ΔΨ loss in CypD^{-/-} hepatocytes (preloaded with TMRM) treated with 10 μM A23187. TMRM fluorescence intensity was monitored by laser scanning confocal microscopy. Representative real-time images (**g**), average TMRM intensity of individual WT (blue) and CypD^{-/-} (pink) cells (**h**), and the half-life time of the fluorescence intensity of individual cells (**i**) are shown. Data shown as mean ± s.d. The arrow in (**h**) indicates the addition of 5 μM CCCP, which completely dissipated ΔΨ. **j–l**, Increased A23187-induced mitochondrial Ca²⁺ uptake by CypD^{-/-} hepatocytes. Cells were loaded with Rhod2-AM and treated with 10 μM A23187. Rhod2 fluorescence intensity was monitored by laser scanning confocal microscopy. Representative images of baseline and peak fluorescence are shown (**j**). Average (**k**) and peak (**l**) fluorescence intensities of cells are shown (mean ± s.d.). In (**l**), hepatocytes were also loaded with Rhod2-AM in the presence of 1 μM Ru360.

death was also considered to represent necrotic cell death. These results indicate that CypD (and the CsA-sensitive mPT) is involved in necrotic cell death induced by reactive oxygen species or Ca^{2+} overload.

Suppression of A23187-induced cell death by CypD deficiency seemed likely to be due to inhibition of the mPT. To determine whether A23187-induced mPT is suppressed by CypD deficiency in cells, we monitored mitochondrial $\Delta\Psi$ and Ca^{2+} accumulation in the presence of A23187. Hepatocytes were loaded with the $\Delta\Psi$ markers tetramethylrhodamine methylester (TMRM) or Mito Tracker Orange CMTM Ros, and treated with A23187, after which the fluorescence intensity was monitored by real-time imaging. Addition of A23187 caused rapid loss of $\Delta\Psi$ in control (wild-type) hepatocytes, whereas CypD-deficient hepatocytes maintained $\Delta\Psi$ for much longer periods of time (Fig. 3g–i and

Supplementary Fig. 5). The addition of CCCP (carbonyl cyanide m-chlorophenyl-hydrazone), a protonophore, completely dissipated $\Delta\Psi$. Mitochondrial accumulation of Ca^{2+} was investigated using Rhod2-AM¹⁶. After A23187 treatment, CypD-deficient hepatocytes showed a rapid increase in Rhod2 fluorescence intensity, but wild-type hepatocytes showed only a marginal increase (Fig. 3j–l). The validity of using Rhod2 as an indicator of mitochondrial Ca^{2+} under these conditions was confirmed using Ru360, an inhibitor of the mitochondrial calcium uniporter (Fig. 3l). These results indicate that compared to control hepatocytes, CypD-deficient hepatocytes absorb a larger amount of cytosolic Ca^{2+} into their mitochondria without loss of mitochondrial $\Delta\Psi$; this is consistent with the results obtained using isolated mitochondria, and suggests that A23187 induces CypD-dependent mPT in cells.

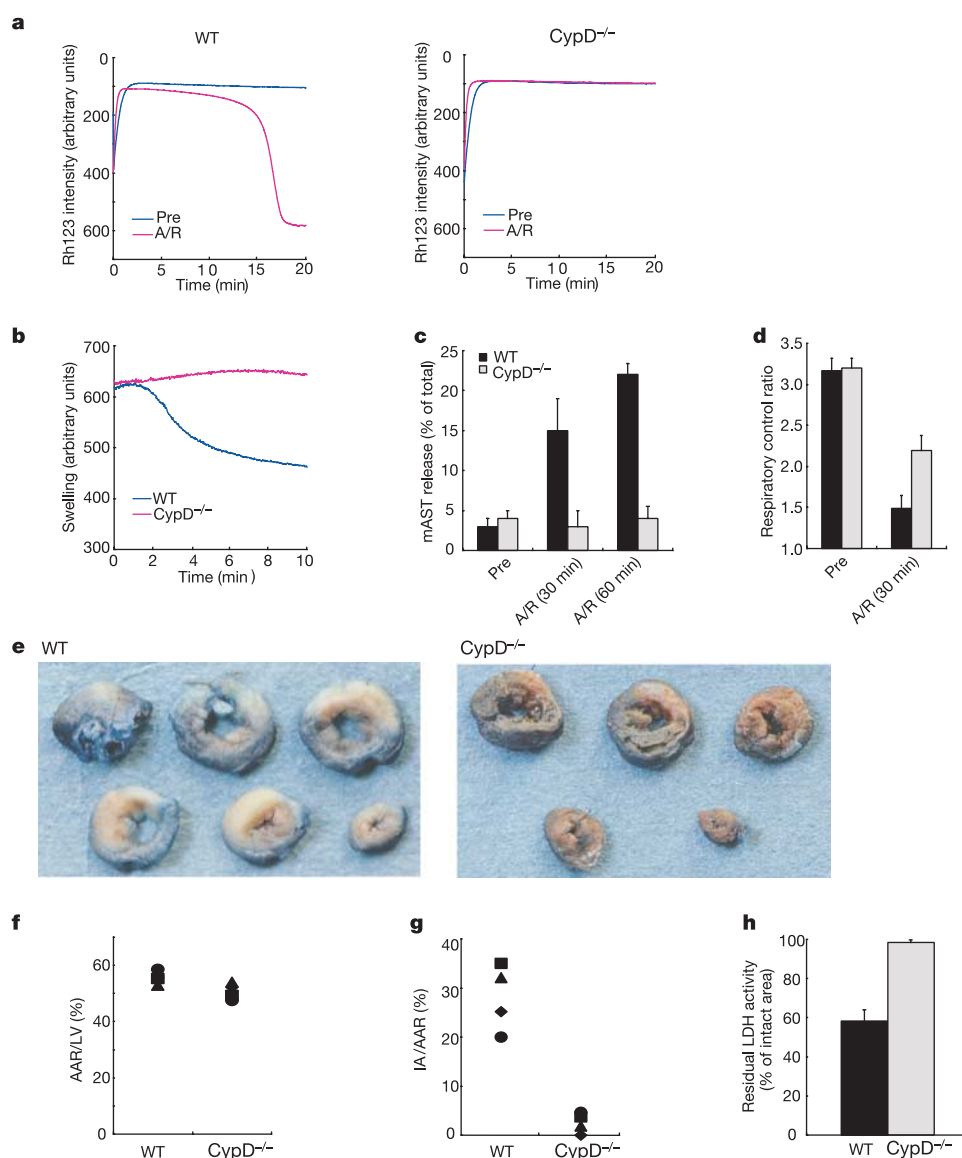


Figure 4 Prevention of cardiac ischaemia/reperfusion injury in CypD^{-/-} mice. **a–d**, Absence of mitochondrial damage induced by anoxia/reoxygenation (A/R) in CypD^{-/-} mitochondria. Isolated mitochondria were treated without (Pre) or with anoxia for 30 min followed by reoxygenation for the indicated times. The $\Delta\Psi$ (**a**), swelling (**b**), mAST release (**c**), and respiratory control ratio (state 3/state 4) (**d**) were measured. Data shown as mean $\pm$ s.e.m. **e–h**, Reduction of cardiac I/R injury in CypD^{-/-} mice.

e, Representative slices of a heart subjected to I/R. The slices were double-stained with Evans blue (blue) and TTC (red). The infarct region was not stained (white). **f**, The area at risk/left ventricle (AAR/LV) and (**g**) the infarct area/area at risk (IA/AAR). Data are shown for four independent experiments. **h**, The ratio of residual LDH activity in the IA/non-ischaemic area. Data shown as mean $\pm$ s.d.

Finally, we investigated the role of CypD in ischaemia/reperfusion (I/R) injury, in which disturbance of Ca^{2+} homeostasis and generation of reactive oxygen species have been implicated¹⁷. Many reports have described a protective effect of CsA against I/R injury^{18–21}. First, we investigated whether CypD-deficient mitochondria showed resistance to anoxia/reoxygenation-induced injury, which simulates I/R-induced injury *in vivo*. Isolated mitochondria from control and CypD-deficient mice were subjected to anoxia for 30 min, followed by reoxygenation. Control mitochondria, but not CypD-deficient mitochondria, showed loss of $\Delta\Psi$ (Fig. 4a), swelling (Fig. 4b), leakage of mitochondrial aspartate aminotransferase (mAST) (Fig. 4c), and a severe decrease in respiratory control rate (Fig. 4d), indicating that CypD-deficient mitochondria are more resistant to anoxia/reoxygenation injury than control mitochondria.

We next examined the effect of CypD on cardiac I/R injury, because the heart has high levels of CypD (see Supplementary Fig. 1d). Several functional parameters assessed by echocardiography showed no differences between the resting hearts of control and CypD-deficient mice (see Supplementary Table). Mice were subjected to 30 min of left coronary artery occlusion followed by 2 h of reperfusion. The size of the area at risk, identified by the absence of Evans blue staining, was not significantly different between control and CypD-deficient hearts (Fig. 4f). In control hearts, I/R injury caused significant necrotic damage, as evidenced by a large area of myocardium that was negative for triphenyltetrazolium chloride (TTC) staining (Fig. 4e, g). In CypD-deficient hearts, however, the infarct area was dramatically reduced (Fig. 4e, g). Consistently, lactate dehydrogenase (LDH) release due to disruption of the plasma membrane was almost completely inhibited in CypD-deficient hearts (Fig. 4h). These results indicate that lack of CypD can markedly reduce cardiac I/R injury.

An increase in the permeability of the outer mitochondrial membrane is central to apoptotic signalling, and is directly regulated by the Bcl-2 family of proteins¹. It has been suggested that the mPT plays a role in apoptotic mitochondrial membrane permeabilization²². However, we show here that cytochrome *c* release induced by Bid and Bax is not blocked by CypD deficiency and that CypD deficiency does not affect many forms of apoptotic cell death, indicating that the CypD-dependent mPT does not play a significant role in apoptosis in general; however, this does not exclude the possibility that certain forms of apoptosis are mediated by the mPT, and thereby inhibited by CsA. On the other hand, CypD deficiency blocks Ca^{2+} -induced and oxidative stress-induced cytochrome *c* release from isolated mitochondria and also prevents necrotic cell death induced by these stimuli; this indicates that the CypD-dependent mPT is a critical event in some forms of cellular necrosis. Notably, overexpressed CypD can induce necrosis²³.

We showed that lack of CypD markedly suppresses cardiac I/R injury (which mimics cardiac infarction). In I/R injury, production of reactive oxygen species and Ca^{2+} overload are known to be key events¹⁷. Our results indicate that CypD-deficient mitochondria can accumulate excess Ca^{2+} without loss of $\Delta\Psi$ and that they also tolerate reactive oxygen species-induced damage. Consistently, CypD-deficient hepatocytes accumulated more Ca^{2+} than wild-type hepatocytes, and were significantly more resistant to A23187- and H_2O_2 -induced death. Thus, the CypD-dependent mPT is a critical event in I/R injury, suggesting that CypD and the mPT may be important therapeutic targets for preventing myocardial infarction. □

Methods

Generation of cyclophilin D-deficient mice

A genomic clone (pKOS 63) containing the CypD locus was isolated from a 129/J mouse genomic library. Lex-1 embryonic stem (ES) cells (derived from the 129SvEvBrd strain) were electroporated with the targeting vector and selected by G418. ES clones with the

targeted CypD allele were screened by polymerase chain reaction (PCR) and Southern blotting analysis. Heterozygous mutant ES cells were injected into C57BL/6J blastocysts. Germline transmission of mutant alleles to F₁ offspring was confirmed by PCR and Southern blotting, and F₂ offspring from heterozygous intercrosses were genotyped by PCR and Southern blotting.

Genotyping by PCR

The CypD locus was genotyped by PCR. The wild-type allele (553 base pairs, bp) was detected using a forward primer (5'-GCAGATCAAGTCCCGACTG-3') and a reverse primer (5'-ACTTGGGAAGCCGAGGTG-3'). To detect the mutant allele (206 bp), a neomycin-specific reverse primer (5'-GCAGCGCATCGCCTTCTATC-3') was used in combination with the wild-type reverse primer described above.

Antibodies

Anti-mouse CypD antibody was produced using a peptide (amino acids 43–57) of CypD. Anti-cytochrome *c*, anti-cyclophilin A and anti-GAPDH antibodies were purchased from Pharmingen, UpState and Biogenesis, respectively.

Mitochondrial biochemical parameters

Mitochondria were isolated from mouse livers as described previously²⁴. In all experiments, except for assessment of respiration, mitochondria were in medium containing 0.3 mM mannitol, 10 mM potassium HEPES (pH 7.4), 0.1% fatty acid-free BSA, 300 μM potassium phosphate (pH 7.4) and 4.2 mM succinate.

Mitochondrial PPLase activity was measured as described previously²⁵, except that the final concentration of the synthetic peptide substrate (succinyl-Ala-Ala-Pro-Phe-4-nitroanilide) was 75 μM . The activity was expressed as K/K_0 , where K and K_0 are the first-order rate constants in the presence and absence of the mitochondrial lysate, respectively. The mitochondrial membrane potential ($\Delta\Psi$) was assessed by measuring the $\Delta\Psi$ -dependent uptake of rhodamine 123 (Rh123, ref. 24). Mitochondrial swelling was monitored by the decrease of 90° light scatter at 520 nm, which was determined using a spectrophotometer (Hitachi F-4500). Mitochondrial respiration was measured with an O₂ electrode (Rank Brothers) in respiration buffer²⁴. The external mitochondrial Ca^{2+} concentration was monitored using a Ca^{2+} -specific electrode (Orion). The amount of intra-mitochondrial Ca^{2+} was measured using $^{45}\text{Ca}^{2+}$. Isolated mitochondria were incubated with $^{45}\text{Ca}^{2+}$ for 25 min, and then the mitochondria (1 mg protein) were centrifuged at 12,000g for 3 min. The pellet and supernatant fractions were collected and their radioactivity measured using a liquid scintillation counter (Wallac 1409). Cytochrome *c* release was determined as previously described¹².

In anoxia/reoxygenation experiments, mitochondria were suspended in a tube and treated with 100% N₂ gas. The gas-saturated tube was sealed with parafilm and left for 30 min, and the mitochondria were then reoxygenated by shaking in a 12-well plate at room temperature for 30 min.

Cell death assay

Primary cultures of MEFs were obtained from CypD-deficient and control littermate embryos at embryonic day 14.5. MEFs immortalized by SV40 T antigen were also used. Hepatocytes were isolated from CypD-deficient mice and their control littermates at 3–4 months of age using the retrograde two-step collagenase perfusion technique²⁶. Cell viability was assessed by propidium iodide (PI) staining, Annexin-V staining or the cell titer blue (CTB) assay. Briefly, cells were stained with 1 μM PI or 1 μM Cy3-conjugated Annexin-V for 5 min at room temperature and were analysed using a flow cytometer (Becton-Dickinson). The CTB assay, which measures the metabolic activity of viable cells, was carried out using Cell Titer Blue reagent (Promega). The nuclear morphology was observed with 10 μM Hoechst 33342 and 1 μM PI staining as described previously²⁷. DEVDase activity (cleavage of the caspase-3 substrate DEVD-amc) was measured as described elsewhere²⁸.

Laser scanning confocal microscopy

Hepatocytes were cultured in covered glass-bottomed 24-well dishes coated with type I collagen (Iwaki glass) for 16–20 h. To monitor mitochondrial $\Delta\Psi$, hepatocytes were loaded with 0.25 μM TMRM (Molecular Probes) as described previously²⁹. To monitor the mitochondrial Ca^{2+} level, hepatocytes were loaded with 5 μM Rhod2-AM (Molecular Probes) in culture medium containing 0.05% Pluronic F-127 (Molecular Probes) and 0.1 mM sulphydrylase for 2 h at 4 °C for mitochondria-selective loading as described previously¹⁶.

Ischaemia/reperfusion experiments

Using CypD-deficient and wild-type littermate mice at 10–12 weeks of age, ischaemia/reperfusion of the heart was performed as described previously³⁰. Briefly, under general anaesthesia with mechanical ventilation, a silk thread (7-0) was passed around the left coronary artery (LCA) about 1 mm distal to the LCA origin to make a snare. After 30 min ligation of the LCA, the snare was released for 2 h. The infarct size was evaluated by double staining using Evans blue dye and triphenyltetrazolium chloride (TTC). The area at risk was defined as the ratio of the area of the ischaemic region to the left ventricular area, and the infarct size was defined as the ratio of the area of the infarct region to that of the ischaemic region.

Statistical evaluation

Statistical evaluation was performed by unpaired *t*-tests. Data are presented as mean $\pm$ s.e.m. or mean $\pm$ s.d.

Received 2 September 2004; accepted 4 January 2005; doi:10.1038/nature03317.

1. Tsujimoto, Y. Cell death regulation by the Bcl-2 protein family in the mitochondria. *J. Cell. Physiol.* **195**, 158–167 (2003).
2. Green, D. R. & Kroemer, G. The pathophysiology of mitochondrial cell death. *Science* **305**, 626–629 (2004).
3. Halestrap, A. P., McStay, G. P. & Clarke, S. J. The permeability transition pore complex: another view. *Biochimie* **84**, 153–166 (2002).
4. Crompton, M. On the involvement of mitochondrial intermembrane junctional complexes in apoptosis. *Curr. Med. Chem.* **10**, 1473–1484 (2003).
5. Kokoszka, J. E. *et al.* The ADP/ATP translocator is not essential for the mitochondrial permeability transition pore. *Nature* **427**, 461–465 (2004).
6. Halestrap, A. P. Mitochondrial permeability: dual role for the ADP/ATP translocator? *Nature* [online] **430**, 983 (2004) (doi:10.1038/nature02816).
7. Galat, A. & Metcalfe, S. M. Peptidylproline cis/trans isomerases. *Prog. Biophys. Mol. Biol.* **63**, 67–118 (1995).
8. Broekemeier, K. M., Dempsey, M. E. & Pfeiffer, D. R. Cyclosporin A is a potent inhibitor of the inner membrane permeability transition in liver mitochondria. *J. Biol. Chem.* **264**, 7826–7830 (1989).
9. He, L. & Lemasters, J. J. Regulated and unregulated mitochondrial permeability transition pores: a new paradigm of pore structure and function? *FEBS Lett.* **512**, 1–7 (2002).
10. Scorrano, L. *et al.* A distinct pathway remodels mitochondrial cristae and mobilizes cytochrome c during apoptosis. *Dev. Cell* **2**, 55–67 (2002).
11. Wei, M. C. *et al.* Proapoptotic BAX and BAK: a requisite gateway to mitochondrial dysfunction and death. *Science* **292**, 727–730 (2001).
12. Narita, M. *et al.* Bax interacts with the permeability transition pore to induce permeability transition and cytochrome c release in isolated mitochondria. *Proc. Natl Acad. Sci. USA* **95**, 14681–14686 (1998).
13. Eskes, R. *et al.* Bax-induced cytochrome C release from mitochondria is independent of the permeability transition pore but highly dependent on Mg^{2+} ions. *J. Cell Biol.* **143**, 217–224 (1998).
14. Finucane, D. M., Bossy-Wetzel, E., Waterhouse, N. J., Cotter, T. G. & Green, D. R. Bax-induced caspase activation and apoptosis via cytochrome c release from mitochondria is inhibitable by Bcl-xL. *J. Biol. Chem.* **274**, 2225–2233 (1999).
15. von Ahsen, O. *et al.* Preservation of mitochondrial structure and function after Bid- or Bax-mediated cytochrome c release. *J. Cell Biol.* **150**, 1027–1036 (2000).
16. Trollinger, D. R., Cascio, W. E. & Lemasters, J. J. Selective loading of Rhod 2 into mitochondria shows mitochondrial Ca^{2+} transients during the contractile cycle in adult rabbit cardiac myocytes. *Biochem. Biophys. Res. Commun.* **236**, 738–742 (1997).
17. Weiss, J. N., Korge, P., Honda, H. M. & Ping, P. Role of the mitochondrial permeability transition in myocardial disease. *Circ. Res.* **93**, 292–301 (2003).
18. Shimizu, S. *et al.* Beneficial effects of cyclosporine on reoxygenation injury in hypoxic rat liver. *Transplantation* **57**, 1562–1566 (1994).
19. Javadov, S. A. *et al.* Ischaemic preconditioning inhibits opening of mitochondrial permeability transition pores in the reperfused rat heart. *J. Physiol. (Lond.)* **549**, 513–524 (2003).
20. Matsumoto, S., Friberg, H., Ferrand-Drake, M. & Wieloch, T. Blockade of the mitochondrial permeability transition pore diminishes infarct size in the rat after transient middle cerebral artery occlusion. *J. Cereb. Blood Flow Metab.* **19**, 736–741 (1999).
21. Khaspekov, L., Friberg, H., Halestrap, A., Viktorov, I. & Wieloch, T. Cyclosporin A and its nonimmunosuppressive analogue N-Me-Val-4-cyclosporin A mitigate glucose/oxygen deprivation-induced damage to rat cultured hippocampal neurons. *Eur. J. Neurosci.* **11**, 3194–3198 (1999).
22. Zamzami, N. & Kroemer, G. The mitochondrion in apoptosis: how Pandora's box opens. *Nature Rev. Mol. Cell Biol.* **2**, 67–71 (2001).
23. Li, Y., Johnson, N., Capano, M., Edwards, M. & Crompton, M. Cyclophilin-D promotes the mitochondrial permeability transition but has opposite effects on apoptosis and necrosis. *Biochem. J.* **383**, 101–109 (2004).
24. Shimizu, S. *et al.* Bcl-2 prevents apoptotic mitochondrial dysfunction by regulating proton flux. *Proc. Natl Acad. Sci. USA* **95**, 1455–1459 (1998).
25. Fischer, G., Wittmann-Liebold, B., Lang, K., Kieffhaber, T. & Schmid, F. X. Cyclophilin and peptidyl-prolyl *cis-trans* isomerase are probably identical proteins. *Nature* **337**, 476–478 (1989).
26. Hatano, E. *et al.* The mitochondrial permeability transition augments Fas-induced apoptosis in mouse hepatocytes. *J. Biol. Chem.* **275**, 11814–11823 (2000).
27. Shinzawa, K. & Tsujimoto, Y. PLA2 activity is required for nuclear shrinkage in caspase-independent cell death. *J. Cell Biol.* **163**, 1219–1230 (2003).
28. Shimizu, S., Eguchi, Y., Kamiike, W., Matsuda, M. & Tsujimoto, Y. Bcl-2 expression prevents activation of the ICE protease cascade. *Oncogene* **12**, 2251–2257 (1996).
29. Byrne, A. M., Lemasters, J. J. & Nieminen, A. L. Contribution of increased mitochondrial free Ca^{2+} to the mitochondrial permeability transition induced by tert-butylhydroperoxide in rat hepatocytes. *Hepatology* **29**, 1523–1531 (1999).
30. Yamashita, N. *et al.* Exercise provides direct biphasic cardioprotection via manganese superoxide dismutase activation. *J. Exp. Med.* **189**, 1699–1706 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to K. Tagawa for helpful discussion and C. Thompson for providing Bak-deficient mice. CypD-deficient mice were developed in collaboration with Lexicon Genetics Incorporated. This study was supported in part by a grant for Scientific Research on Priority Areas, a grant for Center of Excellence Research, a grant for the 21st century COE Program, a grant for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan, and by a grant for Research on Dementia and Fracture from the Ministry of Health, Labour and Welfare of Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.T. (tsujimoto@gene.med.osaka-u.ac.jp).

Loss of cyclophilin D reveals a critical role for mitochondrial permeability transition in cell death

Christopher P. Baines¹, Robert A. Kaiser¹, Nicole H. Purcell¹, N. Scott Blair¹, Hanna Osinska¹, Michael A. Hambleton², Eric W. Brunskill³, M. Richard Sayen⁴, Roberta A. Gottlieb⁴, Gerald W. Dorn II³, Jeffrey Robbins¹ & Jeffery D. Molkentin¹

¹Departments of Pediatrics, ²Molecular Genetics and ³Medicine, University of Cincinnati, Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA
⁴Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California 92037, USA

Mitochondria play a critical role in mediating both apoptotic and necrotic cell death. The mitochondrial permeability transition (mPT) leads to mitochondrial swelling, outer membrane rupture and the release of apoptotic mediators. The mPT pore is thought to consist of the adenine nucleotide translocator, a voltage-dependent anion channel, and cyclophilin D (the *Ppif* gene product), a prolyl isomerase located within the mitochondrial matrix^{1,2}. Here we generated mice lacking *Ppif* and mice overexpressing cyclophilin D in the heart. *Ppif* null mice are protected from ischaemia/reperfusion-induced cell death *in vivo*, whereas cyclophilin D-overexpressing mice show mitochondrial swelling and spontaneous cell death. Mitochondria isolated from the livers, hearts and brains of *Ppif* null mice are resistant to mitochondrial swelling and permeability transition *in vitro*. Moreover, primary hepatocytes and fibroblasts isolated from *Ppif* null mice are largely protected from Ca^{2+} -overload and oxidative stress-induced cell death. However, Bcl-2 family member-induced cell death does not depend on cyclophilin D, and *Ppif* null fibroblasts are not protected from staurosporine or tumour-necrosis factor- α -induced death. Thus, cyclophilin D and the mitochondrial permeability transition are required for mediating Ca^{2+} - and oxidative damage-induced cell death, but not Bcl-2 family member-regulated death.

Disagreement persists as to the function of the mitochondrial permeability transition pore in mediating apoptotic and/or necrotic cell death^{3,4}. To address this issue, we created *Ppif* knockout mice (*Ppif*^{-/-}) using homologous recombination in embryonic stem cells (Fig. 1a). Western blotting of cardiac protein extracts from heterozygous and *Ppif* null mice showed the expected 50% reduction or complete absence of cyclophilin D protein, respectively, without a change in the protein levels of the voltage-dependent anion channel VDAC1 or the adenine nucleotide translocator ANT1/2 (Fig. 1b). Remarkably, mitochondria isolated from heart, liver and brain of *Ppif* null mice were resistant to swelling and permeability transition induced by Ca^{2+} and/or atractyloside compared with mitochondria isolated from wild-type mice (Fig. 1c–f). However, this protective effect was lost at very high Ca^{2+} levels (Supplementary Fig. 1), as previously reported with the mPT inhibitor cyclosporin A (CsA)⁵. *Ppif* null mitochondria also showed a greater overall capacity for Ca^{2+} uptake before undergoing permeability transition (Supplementary Fig. 2). The baseline morphology of mitochondria and cristae organization in *Ppif* null hearts was otherwise unaffected compared to wild-type hearts (Fig. 1g, h). Indeed, the heart, which contains approximately 30% mitochondria by volume, showed no signs of cardiomyopathy in the absence of *Ppif* (Supplementary Fig. 3).

Primary embryonic fibroblast cultures were generated to more carefully evaluate mitochondria and cell death in the absence of *Ppif* (Fig. 2a). Wild-type and *Ppif* null fibroblasts were loaded with tetramethylrhodamine ethyl ester (TMRE) to measure inner mitochondrial membrane potential ($\Delta\Psi_m$) and permeability

transition (Supplementary Fig. 4). Wild-type and null fibroblasts infected with a control adenovirus expressing β -galactosidase showed robust TMRE staining of all mitochondria (Fig. 2b). However, *Ppif* null fibroblasts were resistant to loss of $\Delta\Psi_m$ following H_2O_2 treatment compared with wild-type fibroblasts (Fig. 2b). Protection from H_2O_2 -induced dissipation of $\Delta\Psi_m$ in *Ppif* null fibroblasts was reversed by adenoviral-mediated gene transfer of cyclophilin D, but not a prolyl isomerase-deficient mutant of cyclophilin D (Fig. 2b). Fluorescence-activated cell sorting (FACS) showed a dramatic reduction in TMRE fluorescence following H_2O_2 treatment in wild-type but not in *Ppif* null fibroblasts (Fig. 2c). Wild-type and *Ppif* null fibroblasts were also loaded with calcein-AM in the presence of $CoCl_2$, and again the *Ppif* null cells showed resistance to H_2O_2 -induced mPT (Supplementary Fig. 4). *Ppif* null fibroblasts were also protected from H_2O_2 -induced cell death as assessed by propidium iodide (PI), TUNEL and annexin V staining (Fig. 2d, e and Supplementary Fig. 4). Re-introduction of cyclophilin D, but not a prolyl isomerase-deficient mutant of cyclophilin D, restored efficient cell death in *Ppif* null fibroblasts (Fig. 2d, e). Cultures of primary hepatocytes were generated from wild-type and *Ppif* null mice, and the latter were again protected from oxidative stress-induced cell death, similar to the *Ppif* null fibroblasts (Fig. 2f). *Ppif* null fibroblasts were also protected from cell death induced by

Ca^{2+} overload following thapsigargin or ionomycin treatment, but not from cell death induced by staurosporine or tumour necrosis factor- α (TNF- α , Fig. 2g).

Wild-type and *Ppif* null mice were also subjected to cardiac ischaemia/reperfusion injury, and a 40% reduction in infarction area was observed in the *Ppif* null mice, nearly identical to the level of protection observed using CsA in wild-type mice (Fig. 3a). To investigate the potential gain-of-function effects associated with cyclophilin D on mPT, cardiac-specific transgenic mice overexpressing cyclophilin D were generated (Fig. 3b). The lowest-expressing transgenic line (line 26.2) and the highest-expressing line (line 25.4) were examined further, and showed alterations in heart mitochondrial architecture and swelling as well as loss of cristae at two months of age (Fig. 3c), although mitochondrial number was not altered (data not shown). Mitochondria isolated from the hearts of low- and high-expressing transgenic mice showed pronounced swelling at baseline that was alleviated by polyethylene glycol (PEG)-induced shrinkage in a CsA-sensitive manner (Fig. 3d, e). Cyclophilin D transgenic mice also showed increased TUNEL staining in the heart (Fig. 3f), increased cytochrome *c* release, increased caspase 9 cleavage, cardiac hypertrophy, and a reduction in cardiac functional performance over time (Supplementary Fig. 5).

The intrinsic and extrinsic pathways regulating cell death

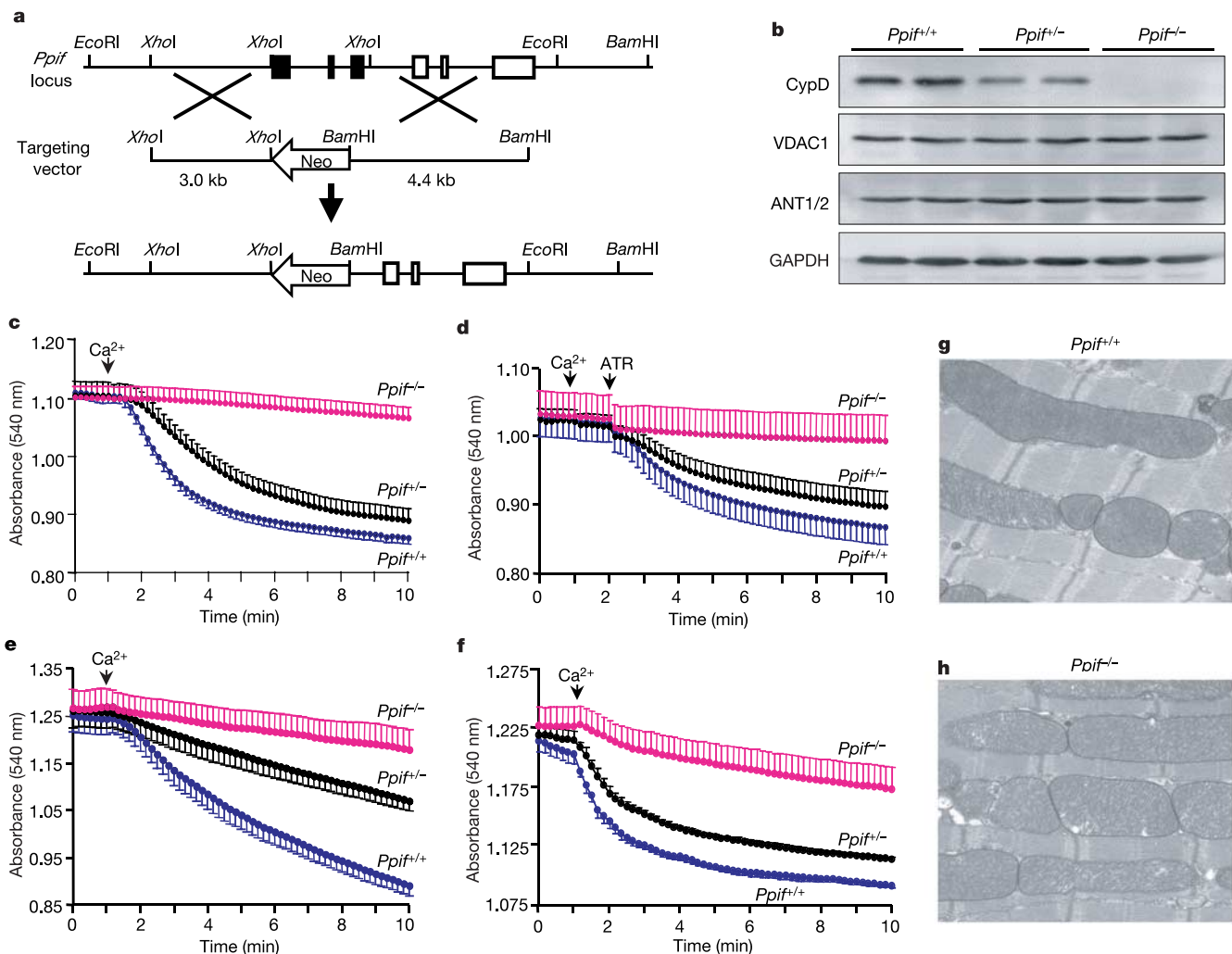


Figure 1 Generation of *Ppif*^{-/-} mice and assessment of mitochondrial pore function. **a**, Map of the wild-type *Ppif* locus, the targeting vector and the mutant *Ppif* locus after homologous recombination. **b**, Western blots for Cyclophilin D (CypD), VDAC1, and ANT1/2 in *Ppif*^{+/+}, *Ppif*^{+/-}, and *Ppif*^{-/-} mouse hearts. GAPDH was used as a loading control. **c**, **d**, Swelling in isolated cardiac mitochondria induced by 250 μ M Ca^{2+} (**c**) or

200 μ M attractylolide (ATR) primed with 25 μ M Ca^{2+} (**d**). **e**, **f**, Ca^{2+} -induced (250 μ M) swelling in isolated liver (**e**) and brain (**f**) mitochondria. All error bars show s.e.m. **g**, **h**, Transmission electron micrographs (original magnification $\times 20,000$) in *Ppif*^{+/+} (**g**) and *Ppif*^{-/-} (**h**) hearts.

converge on the mitochondria, in part through the action of the pro-death Bcl-2 family members Bax, Bak and Bid. Indeed, Bax and Bid have both been shown to regulate cytochrome *c* release and/or cell death in communication with mPT^{6–11}, but others have demonstrated that Bax and Bid promote cytochrome *c* release independently of the mPT^{12–14}. Here we generated recombinant Bax and a truncated form of Bid (tBid) in order to assess cytochrome *c* release from purified wild-type or *Ppif* null mitochondria. Bax and tBid caused a similar time-dependent profile of cytochrome *c* release into the supernatant, with a corresponding depletion in the mitochondrial protein fraction in both wild-type and *Ppif* null mitochondria; however, 180 min incubation with non-specific control proteins had no effect (Fig. 4a, b). Cytochrome oxidase IV (COX-IV) was used as a control for the purity of the mitochondrial and supernatant protein fractions (Fig. 4a, b). Similar

results were observed in a reaction buffer containing Mg²⁺ (data not shown). Although Bax and tBid induced equivalent release of cytochrome *c* in wild-type and *Ppif* null fibroblasts, *Ppif* null mitochondria were substantially more resistant to Ca²⁺-stimulated cytochrome *c* release than wild-type mitochondria (Fig. 4c). Adenoviral-mediated overexpression of Bax (Fig. 4d) resulted in equivalent cell death, caspase 3 cleavage and caspase 9 cleavage in wild-type and *Ppif* null fibroblasts (Fig. 4e, f). Thus, Bax and tBid can induce cytochrome *c* release and initiate mitochondrial-dependent cell death through a mitochondrial pore-independent mechanism; this is consistent with the lack of cell death protection observed in *Ppif* null fibroblasts in response to staurosporine or TNF- α treatment.

Although numerous studies using the cyclophilin D inhibitors CsA or sanglifehrin A, and the ANT inhibitor bongkreikic acid, have suggested the importance of mPT in facilitating cell death^{1,2,15–21}, the

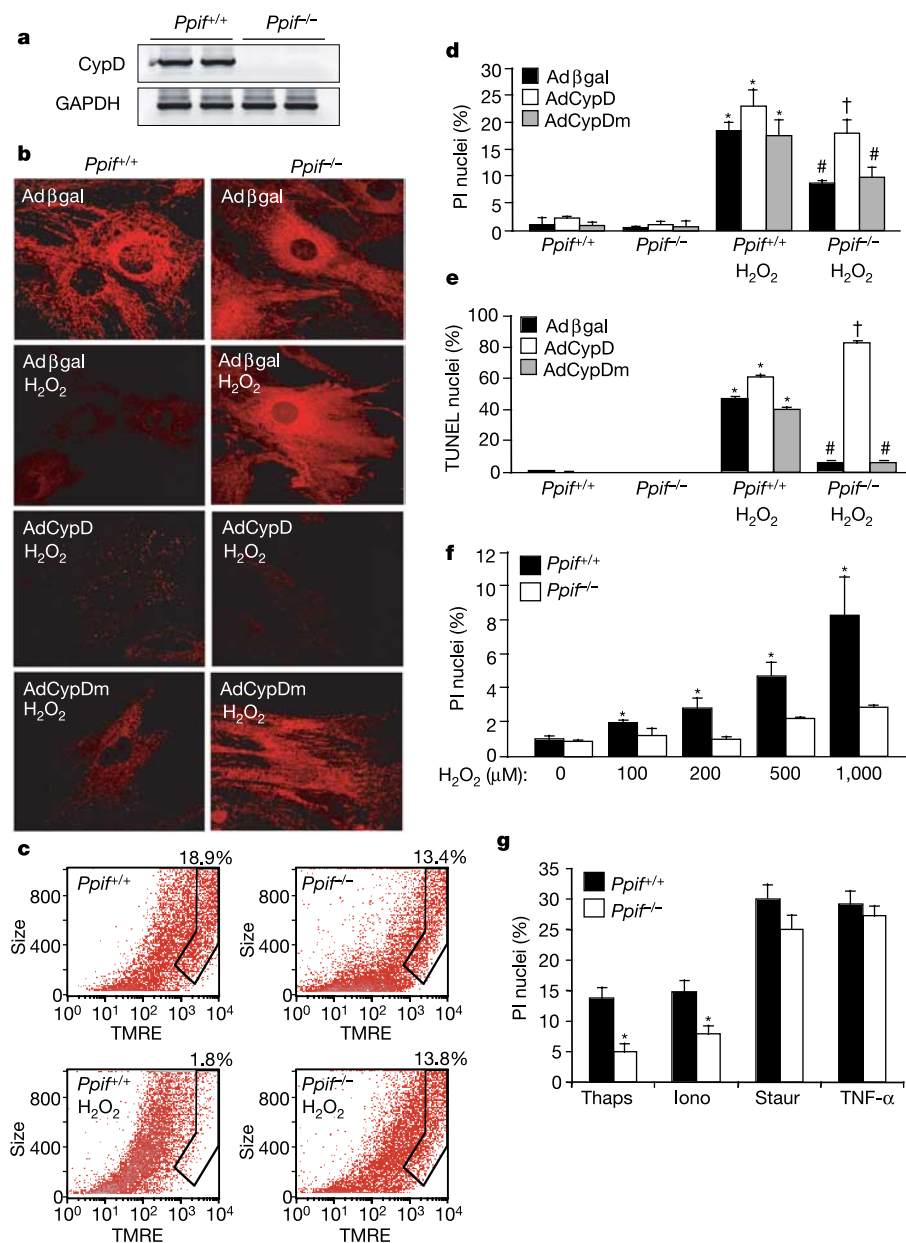


Figure 2 H₂O₂-induced mitochondrial dysfunction and death in *Ppif*^{-/-} cells. **a**, Polymerase chain reaction with reverse transcription (RT-PCR) of cyclophilin D (CypD) and GAPDH messenger RNA from *Ppif*^{+/+} and *Ppif*^{-/-} MEFs. **b**, Confocal microscopy of TMRE in *Ppif*^{+/+} and *Ppif*^{-/-} MEFs infected with the indicated adenoviruses. **c**, Cultured MEFs were treated and stained as in **b**, and subjected to FACS analysis. **d**, **e**, PI staining (**d**) and TUNEL staining (**e**) in infected MEFs. Cell death was induced with

500 μ M H₂O₂ for 6 h. **f**, PI staining in *Ppif*^{+/+} and *Ppif*^{-/-} hepatocytes treated with increasing H₂O₂ concentrations for 6 h. In **d–f**, asterisk denotes $P < 0.05$ versus untreated control; hash denotes $P < 0.05$ versus *Ppif*^{+/+} H₂O₂; cross denotes $P < 0.05$ versus Ad β Gal-infected *Ppif*^{-/-} H₂O₂. **g**, PI staining in *Ppif*^{+/+} and *Ppif*^{-/-} MEFs treated with 20 μ M thapsigargin (Thaps), 10 μ M ionomycin (Iono), 300 nM staurosporine (Staur) or 100 ng ml⁻¹ TNF- α for 24 h. All error bars show s.e.m.

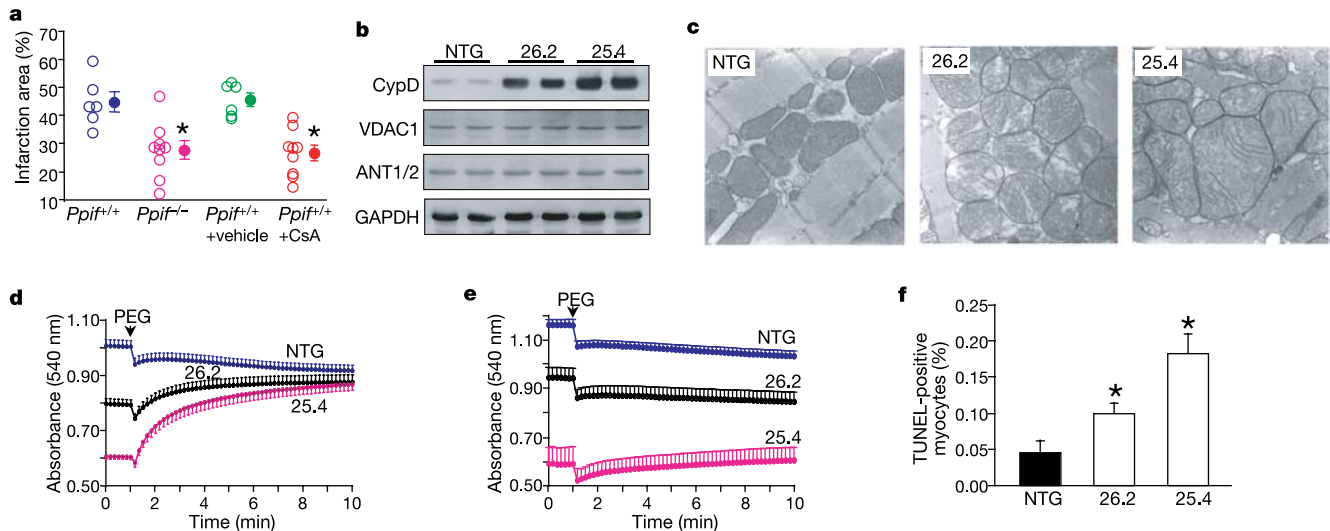


Figure 3 Cardiomyocyte death in *Ppif*^{-/-} mice and CypD transgenic mice. **a**, Myocardial infarction injury area in *Ppif*^{+/+} and *Ppif*^{-/-} mice subjected to 60 min ischaemia and 24 h reperfusion, or *Ppif*^{+/+} mice after treatment with vehicle (veh) or CsA (10 mg kg⁻¹ intraperitoneally every 12 h starting 36 h before infarction). Asterisk denotes *P* < 0.05 versus *Ppif*^{+/+}; error bars show s.e.m. **b**, Western blotting for CypD, VDAC1, ANT1/2 and

GAPDH from hearts of non-transgenic (NTG) and cyclophilin D transgenic mice. **c**, Transmission electron micrographs of mouse hearts. **d**, **e**, PEG-induced shrinking of isolated cardiac mitochondria left untreated (**d**) or pretreated with 5 μM CsA (**e**). **f**, Quantification of TUNEL staining in cardiac sections (4 hearts per group). Asterisk denotes *P* < 0.05 versus NTG; all error bars show s.e.m.

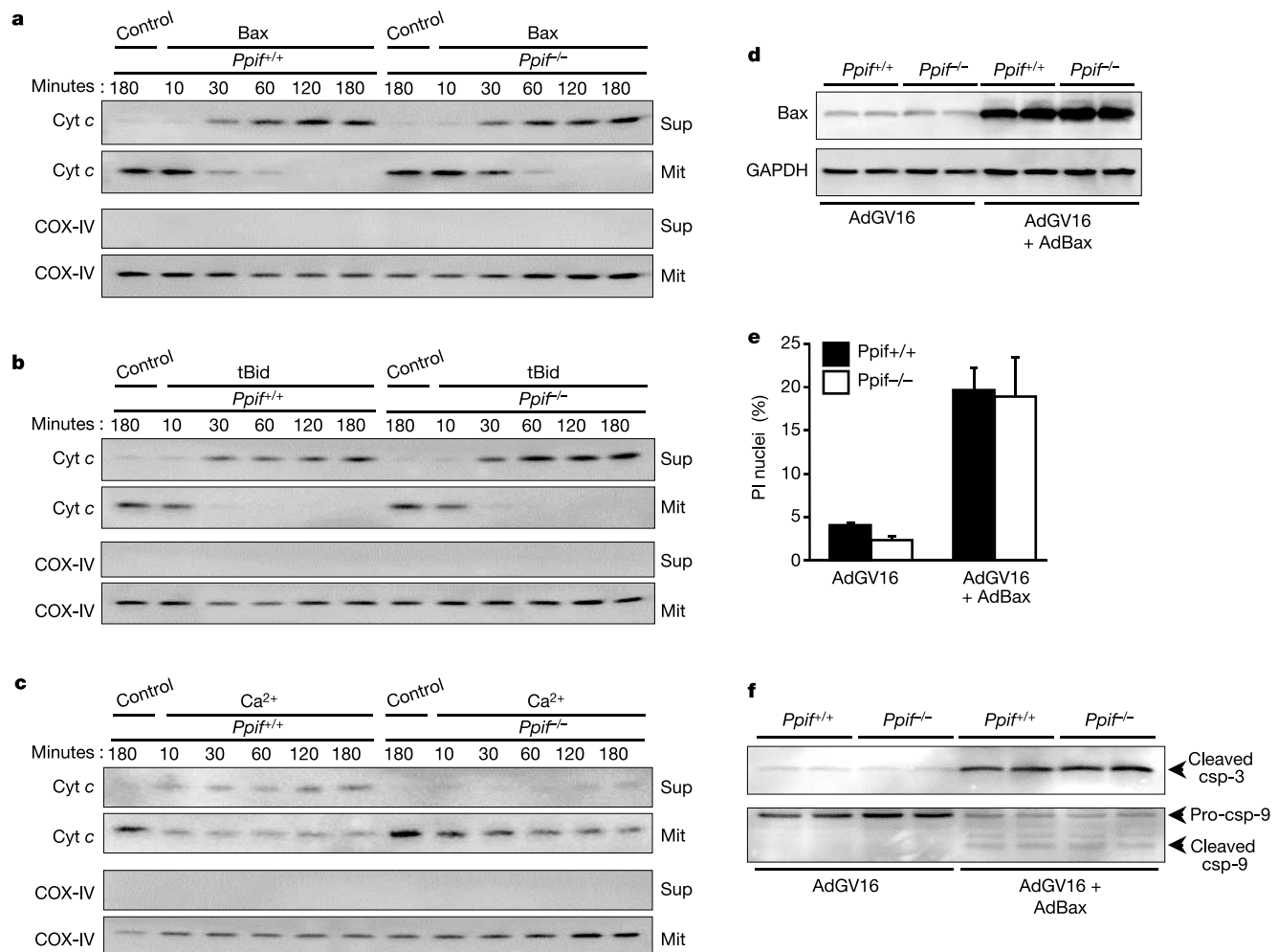


Figure 4 Cytochrome *c* release and cell death induced by Bax, tBid and Ca²⁺. **a–c**, Isolated *Ppif*^{+/+} and *Ppif*^{-/-} liver mitochondria were incubated with GST-Bax (**a**), tBid (**b**), or 100 μM Ca²⁺ (**c**) and the resultant pellet (Mit) and supernatant (Sup) were subjected to western blotting for cytochrome *c* and COX-IV (control). Similar results were observed in a total of 4 independent experiments. **d**, Western blotting for Bax and GAPDH

in *Ppif*^{+/+} and *Ppif*^{-/-} MEFs infected with adenoviruses encoding an inducible Bax (AdBax). **e**, PI staining in AdBax-infected *Ppif*^{+/+} and *Ppif*^{-/-} MEFs (*n* = 4). Error bars show s.e.m. **f**, Western blotting for cleaved caspase-3 and caspase-9 in AdBax-infected *Ppif*^{+/+} and *Ppif*^{-/-} MEFs.

role of mPT as a cell death mediator has been questioned by the observation that liver mitochondria lacking ANT are still capable of permeability transition⁴. Indeed, hepatocytes deficient in *Ant1* (*Slc25a4*) and *Ant2* (*Slc25a5*) showed enhanced Ca^{2+} -induced mitochondrial swelling and cell death. However, swelling in *Ant1/2* null mitochondria was still sensitive to CsA, suggesting that cyclophilin D was still a critical component of the pore. It has been suggested that cyclophilin D might potentially regulate inner and outer mitochondrial membrane contact points independently of ANT⁶; others have proposed the accumulation of unfolded integral membrane proteins that can form chaperone-dependent, Ca^{2+} -regulated pores⁵.

Although loss of *Ppif* protected against Ca^{2+} - and oxidative stress-induced death, cyclophilin D-regulated pore formation cannot regulate all forms of mitochondrial-driven cell death. For example, *Ppif* null fibroblasts were not protected from staurosporine- or TNF- α -mediated cell death, or death induced by Bax. Somewhat consistent with these data, cyclophilin D overexpression in B50 neuronal cells promoted mPT and necrotic cell death, although it inhibited apoptotic cell death²². Cyclophilin D overexpression was also reported to delay staurosporine-induced apoptosis in rat C6 glioma cells in culture, but the necrotic response was not analysed²³.

The mPT has been proposed to either induce or increase the efficiency of cell death by promoting greater rupture of the outer mitochondrial membrane through swelling and enhanced cytochrome *c* release. However, discrete openings in the outer mitochondrial membrane associated with the activities of Bax or tBid, for example, may not necessarily affect inner mitochondrial membrane morphology, swelling or $\Delta\Psi_m$ (refs 12–14). In contrast, Bax/Bak/tBid and their evoking stimuli have been proposed to mediate cytochrome *c* release or cell death through the mitochondrial pore^{7–11}. Our results suggest that mPT does serve as a requisite mediator of Ca^{2+} - and oxidative stress-induced cell death. However, depending on the cell-type and/or inducing stimuli employed, mPT has either no role or only a secondary role in modulating Bax/Bak/tBid-dependent cell death. □

Methods

Generation of *Ppif* null mice and cyclophilin D transgenic mice

The *Ppif*-targeting vector consisted of a 3.0 kilobase (kb) genomic fragment, the neomycin-resistance gene (*neo*) and a 4.4 kb genomic fragment, which replaced the first three coding exons of the *Ppif* gene following homologous recombination (Fig. 1a). Embryonic stem cell clones (Sv129 strain) were used to generate chimaeric mice and eventually *Ppif* null mice (*Ppif* heterozygous matings produced a mendelian distribution of genotypes). The full-length mouse cyclophilin D complementary DNA was subcloned into a murine α -myosin heavy chain promoter expression vector to generate transgenic mice (FVB/N strain).

Histological analyses

For mitochondrial morphological assessment, hearts were fixed in glutaraldehyde and cacodylate, embedded in epoxy resin and sectioned, counterstained with uranyl acetate and lead citrate, and subjected to transmission electron microscopy. Myocardial ischaemia/reperfusion injury and histological analysis of infarction area were described previously²⁴.

Mitochondrial isolation and analyses

Mouse heart, liver and brain mitochondria were isolated by homogenization followed by differential centrifugation²⁵. Pore opening was determined by mitochondrial swelling and shrinking¹⁶, where the light-scattering of 250 μg of mitochondria in a 1 ml volume was measured at 540 nm. All swelling assays or cytochrome *c* release assays were performed in 120 mM KCl, 10 mM Tris pH 7.6 and 5 mM KH_2PO_4 . Swelling was induced by either 250 μM CaCl_2 or 25 μM CaCl_2 plus 200 μM atractyloside, and shrinking was induced by 5% PEG-3350 (w/v). For cytochrome *c* release, 40 μg of liver mitochondria in a 60 μl volume was incubated at 30 °C with 100 μM CaCl_2 , 1 μg recombinant GST-Bax or 0.5 μg recombinant tBid.

Tissue culture

Mouse embryonic fibroblasts (MEFs) were isolated from day 15.5–16.5 embryos. MEFs were assayed for cell death by both PI exclusion and TUNEL staining (Roche). $\Delta\Psi_m$ and pore opening were measured by TMRE and calcein-AM/ CaCl_2 staining, respectively²⁶. Primary cultured hepatocytes were isolated by *in situ* collagenase (Sigma) perfusion and plated on collagen-coated dishes.

Adenovirus construction

Replication-deficient adenoviruses for β -galactosidase (Ad βgal), wild-type mouse cyclophilin D (AdCypD) or a R96G isomerase-deficient mutant of cyclophilin D

(AdCypDm) were generated using the AdEasy adenoviral system (Stratagene). The binary Bax adenoviral expression system (a gift from B. Fang) has been previously described²⁷. MEFs were typically infected with adenovirus at a multiplicity of infection of 100 plaque-forming units for 2 h at 37 °C. The cells were then cultured for a further 24 h in virus-free media before treatment.

Received 4 November 2004; accepted 7 February 2005; doi:10.1038/nature03434.

- Halestrap, A. P., McStay, G. P. & Clarke, S. J. The permeability transition pore complex: another view. *Biochimie* **84**, 153–166 (2002).
- Crompton, M. On the involvement of mitochondrial intermembrane junctional complexes in apoptosis. *Curr. Med. Chem.* **10**, 1473–1484 (2003).
- Halestrap, A. P. Mitochondrial permeability: dual role of the ADP/ATP translocator? *Nature* **430**, doi:10.1038/nature02816 (2004).
- Kokoszka, J. E. *et al.* The ADP/ATP translocator is not essential for the mitochondrial permeability transition pore. *Nature* **427**, 461–465 (2004).
- He, L. & Lemasters, J. J. Regulated and unregulated mitochondrial permeability transition pores: a new paradigm of pore structure and function? *FEBS Lett.* **512**, 1–7 (2002).
- Scorrano, L. *et al.* A distinct pathway remodels mitochondrial cristae and mobilizes cytochrome *c* during apoptosis. *Dev. Cell* **2**, 55–67 (2002).
- Zamzami, N. *et al.* Bid acts on the permeability transition pore complex to induce apoptosis. *Oncogene* **19**, 6342–6350 (2000).
- Marzo, I. *et al.* Bax and adenine nucleotide translocator cooperate in the mitochondrial control of apoptosis. *Science* **281**, 2027–2031 (1998).
- Pastorino, J. G., Chen, S. T., Tafani, M., Snyder, J. W. & Farber, J. L. The overexpression of Bax produces cell death upon induction of the mitochondrial permeability transition. *J. Biol. Chem.* **273**, 7770–7775 (1998).
- Narita, M. *et al.* Bax interacts with the permeability transition pore to induce permeability transition and cytochrome *c* release in isolated mitochondria. *Proc. Natl Acad. Sci. USA* **95**, 14681–14686 (1998).
- Shimizu, S., Narita, M. & Tsujimoto, Y. Bcl-2 family proteins regulate the release of apoptogenic cytochrome *c* by the mitochondrial channel VDAC. *Nature* **399**, 483–487 (1999).
- Esques, R. *et al.* Bax-induced cytochrome *c* release from mitochondria is independent of the permeability transition pore but highly dependent on Mg^{2+} ions. *J. Cell Biol.* **143**, 217–224 (1998).
- von Ahsen, O. *et al.* Preservation of mitochondrial structure and function after Bid- or Bax-mediated cytochrome *c* release. *J. Cell Biol.* **150**, 1027–1036 (2000).
- Bossy-Wetzel, E., Newmeyer, D. D. & Green, D. R. Mitochondrial cytochrome *c* release in apoptosis occurs upstream of DEVD-specific caspase activation and independently of mitochondrial transmembrane depolarization. *EMBO J.* **17**, 37–49 (1998).
- Griffiths, E. J. & Halestrap, A. P. Protection by Cyclosporin A of ischemia/reperfusion-induced damage in isolated rat hearts. *J. Mol. Cell. Cardiol.* **25**, 1461–1469 (1993).
- Clarke, S. J., McStay, G. P. & Halestrap, A. P. Sanglifehrin A acts as a potent inhibitor of the mitochondrial permeability transition and reperfusion injury of the heart by binding to cyclophilin-D at a different site from cyclosporin A. *J. Biol. Chem.* **277**, 34793–34799 (2002).
- Adams, J. W. *et al.* Cardiomyocyte apoptosis induced by G α_q signaling is mediated by permeability transition pore formation and activation of the mitochondrial death pathway. *Circ. Res.* **87**, 1180–1187 (2000).
- Imberti, R., Nieminen, A. L., Herman, B. & Lemasters, J. J. Mitochondrial and glycolytic dysfunction in lethal injury to hepatocytes by t-butylhydroperoxide: protection by fructose, cyclosporin A and trifluoperazine. *J. Pharmacol. Exp. Ther.* **265**, 392–400 (1993).
- Griffiths, E. J. & Halestrap, A. P. Mitochondrial non-specific pores remain closed during cardiac ischaemia, but open upon reperfusion. *Biochem. J.* **307**, 93–98 (1995).
- Nieminen, A. L., Saylor, A. K., Tesfai, S. A., Herman, B. & Lemasters, J. J. Contribution of the mitochondrial permeability transition to lethal injury after exposure of hepatocytes to t-butylhydroperoxide. *Biochem. J.* **307**, 99–106 (1995).
- Qian, T., Nieminen, A. L., Herman, B. & Lemasters, J. J. Mitochondrial permeability transition in pH-dependent reperfusion injury to rat hepatocytes. *Am. J. Physiol.* **273**, C1783–C1792 (1997).
- Li, Y., Johnson, N., Capano, M., Edwards, M. & Crompton, M. Cyclophilin-D promotes the mitochondrial permeability transition but has opposite effects on apoptosis and necrosis. *Biochem. J.* **383**, 101–109 (2004).
- Lin, D. T. & Lechleiter, J. D. Mitochondrial targeted cyclophilin D protects cells from cell death by peptidyl prolyl isomerization. *J. Biol. Chem.* **277**, 31134–31141 (2002).
- Kaiser, R. A. *et al.* Targeted inhibition of p38 mitogen-activated protein kinase antagonizes cardiac injury and cell death following ischemia-reperfusion *in vivo*. *J. Biol. Chem.* **279**, 15524–15530 (2004).
- Pallotti, F. & Lenaz, G. Isolation and subfractionation of mitochondria from animal cells and tissue culture lines. *Methods Cell Biol.* **65**, 1–35 (2001).
- Petronilli, V. *et al.* The mitochondrial permeability transition, release of cytochrome *c* and cell death. Correlation with the duration of pore openings *in situ*. *J. Biol. Chem.* **276**, 12030–12034 (2001).
- Kagawa, S. *et al.* A binary adenoviral vector system for expressing high levels of the proapoptotic gene *bax*. *Gene Ther.* **7**, 75–79 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements @We would like to thank R. N. Kitsis for critical evaluation of this manuscript, M. E. Rothenberg, E. B. Brandt and S. P. Hogan for assistance with FACS analysis, M. B. Cohen and C. Liu for assistance in generating primary hepatocytes, and N. Zimmerman for help with mitochondrial Ca^{2+} uptake assays. This work was supported by Grants from the National Institutes of Health (J.D.M., J.R., G.W.D. and R.A.G.), by an American Heart Association Established Investigator Grant (J.D.M.) and a National Institute of Health NRSA service award (R.A.K. and N.H.P.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.D.M. (jeff.molkentin@cchmc.org).

Structural insights into mRNA recognition from a PIWI domain–siRNA guide complex

James S. Parker, S. Mark Roe & David Barford

Section of Structural Biology, Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

RNA interference and related RNA silencing phenomena use short antisense guide RNA molecules to repress the expression of target genes^{1,2}. Argonaute proteins³, containing amino-terminal PAZ (for PIWI/Argonaute/Zwille) domains and carboxy-terminal PIWI domains, are core components of these mechanisms. Here we show the crystal structure of a Piwi protein from *Archaeoglobus fulgidus* (AfPiwi) in complex with a small interfering RNA (siRNA)-like duplex, which mimics the 5' end of a guide RNA strand bound to an overhanging target messenger

RNA. The structure contains a highly conserved metal-binding site that anchors the 5' nucleotide of the guide RNA. The first base pair of the duplex is unwound, separating the 5' nucleotide of the guide from the complementary nucleotide on the target strand, which exits with the 3' overhang through a short channel. The remaining base-paired nucleotides assume an A-form helix, accommodated within a channel in the PIWI domain, which can be extended to place the scissile phosphate of the target strand adjacent to the putative slicer catalytic site. This study provides insights into mechanisms of target mRNA recognition and cleavage by an Argonaute–siRNA guide complex.

RNA silencing is mediated by the effector complexes RISC (for RNA-induced silencing complex) and RITS (for RNA-induced initiation of transcriptional gene silencing), functioning in post-transcriptional and transcriptional silencing, respectively^{4–6}. Argonautes, comprising the Ago and Piwi subfamilies³, are the only proteins common to these complexes. Recent structural and biochemical studies implicate the PAZ domain in the recognition of an RNA 3' overhang^{7–10}. The PIWI domain is believed to incorporate the slicer catalytic site, which is responsible for the cleavage of target mRNA during RNA interference^{11–13}. However, these results

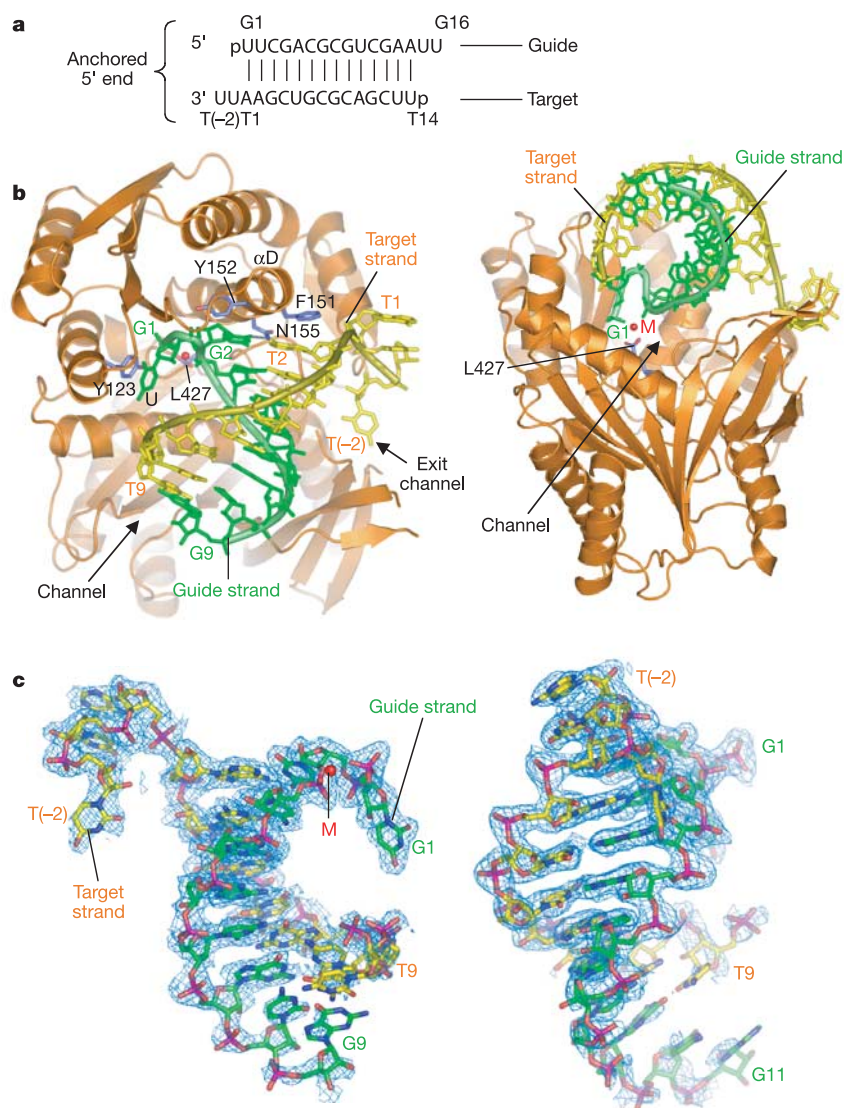


Figure 1 General features of AfPiwi–siRNA interactions and distortion of the 5' terminus of the RNA duplex. **a**, Diagram of the RNA duplex co-crystallized with AfPiwi, indicating the nomenclature for guide and target strands. **b**, Orthogonal views of the AfPiwi–RNA complex. The divalent metal in the G1-binding pocket is shown as a red sphere (M).

c, Views of the RNA 2F_o–F_c electron density map contoured at 1σ. Left, view of the RNA as seen by the protein. Right, orthogonal view. G1 to G9 (G1 to G11 right) and T(-2) to T9 are placed in electron density. The figure was generated with PYMOL (<http://pymol.sourceforge.net>).

are insufficient to explain the requirement for a 5' phosphate group on the first nucleotide of the guide strand^{13–16}, and the observation that the identity of the scissile phosphate in the target mRNA is measured relative to the 5' end of the guide strand, irrespective of modification or the length (within limits) of the 3' end^{15–18}.

AfPiwi, an isolated archaeal PIWI domain-containing protein, which serves as a model for understanding eukaryotic Argonaute PIWI domains, forms a single discrete complex with a 16-nucleotide siRNA-like duplex¹². With the aim of providing a molecular picture of PIWI domain–RNA interactions, we have crystallized and determined the structure of this complex at 2.2 Å resolution. We assign the two RNA strands as guide and target, denoting the guide nucleotides G1 to G16 (5' → 3') and complementary nucleotides T1 to T14 (Fig. 1a). The 5' end of the duplex refers to the 5' end of the guide.

AfPiwi binds to the 5' end of the siRNA-like duplex, engaging the RNA in an L-shaped channel (Fig. 1b). The RNA electron density is strongest at this end, allowing us to model nucleotides G1–G9 of the guide strand and T(–2)–T9 of the target (Fig. 1c). Nucleotides 2–9 on each strand base pair, adopting an A-form helix that extends along the channel (Fig. 1b and Supplementary Fig. 1). Strikingly, the

5' end of the RNA duplex is significantly distorted, resulting in the unwinding of G1 (uridine) and T1 (adenosine). G1, which carries the obligatory 5' phosphate of the guide, is displaced into a pocket surrounding the C terminus of the PIWI domain (G1-binding pocket). T1 passes through the short 'exit' portion of the channel, allowing the two 3' overhanging nucleotides, T(–1) and T(–2), to fold back along the protein surface. G1 and T1 are divided by the αD helix, which contributes two aromatic residues (Phe 151 and Tyr 152) capping the first 5' base pair of the RNA helix (Fig. 1b).

The G1-binding pocket is part of a highly conserved region in the PIWI domain, designated CRI (conserved region I)¹². A prominent structural feature is a divalent metal ion coordinated to the carboxylate group of the C-terminal residue, Leu 427 (Fig. 2a). This metal ion is crucial in the distortion and anchoring of the guide RNA, contacting the phosphates of G1 and G3. Without exception, all Argonaute sequences share three conserved hydrophobic residues at their extreme C terminus (Supplementary Fig. 2), indicating that an exposed carboxylate is an invariant structural feature.

Within the G1-binding pocket, the side chains of four invariant residues in Argonaute proteins (Tyr 123, Lys 127, Gln 137 and Lys 163; Supplementary Fig. 2) contact the 5' phosphate group of G1 (Fig. 2a). Tyr 123 further stabilizes the flipped conformation of G1 by stacking with the unpaired uracil base. In addition to interactions with these invariant side chains and the metal ion, the 5' phosphate contacts the backbone amide of Phe 138, fixed in position by the adjacent Gln 137. No other amino acids contact G1. Only five other residues are invariant in the PIWI domains of Argonaute proteins (Supplementary Fig. 2). Four of these seem to maintain the PIWI fold, and the fifth is a putative slicer catalytic site residue^{11,12}. The G1-binding pocket is therefore the most conserved structural feature of the PIWI domain. Consistent with our structural data is the observation that mutation of this pocket significantly decreased the affinity of AfPiwi for the siRNA-like duplex¹².

Interactions between AfPiwi and the guide are also mediated by the phosphate groups of G2–G5 (Fig. 2). A second key anchor site is the phosphate group of G3. In addition to its coordination by the metal, the phosphate interacts with Gln 159 and Arg 380, residues conserved in the Piwi and Ago subfamilies, respectively (Supplementary Fig. 2). The phosphate of G2 contacts Tyr 152, conserved as Tyr or Thr in the Ago subfamily, and the main chain amide of Arg 140. The phosphate of G4 also contacts Arg 380, and the phosphate of G5 interacts with Arg 385, through a water molecule. The only protein–base contact occurs between Asn 155, conserved in the Ago subfamily, and the 2' oxygen of the G2 uracil. Hydrogen bond acceptors occupy equivalent positions in cytosine and purines. Therefore, as expected, none of the interactions provide RNA sequence specificity. Interactions with G1–G5 account for all conserved residues of CRI (Fig. 3b), strongly indicating a conserved mode of RNA 5'-end binding within the Argonaute family. In contrast to the extensive interactions between the PIWI domain and the guide, few contacts are formed to the target strand, which is consistent with its requirement to dissociate after mRNA cleavage^{19,20}.

With G1 distorted into the conserved binding pocket, the corresponding base pair (G1–T1) is unwound (Fig. 1b). The finding that a single guide–target mismatch at this position can enhance slicer activity¹⁹ demonstrates that the AfPiwi–RNA complex corresponds to a cleavage-competent conformation. It also suggests an equilibrium between the preferred 'active' RNA duplex conformation that we observe and an 'inhibitory' conformation in which G1 pairs with T1. The mode of guide RNA binding to the PIWI domain—a highly ordered anchored 5' region, with bases exposed apart from G1—is consistent with the observation that for animal microRNAs, only nucleotides 2–8 of the microRNA guide (the 'seed') are critical for target recognition and pairing^{21–24}. Similarly, in RISC-mediated mRNA cleavage, bases at the 5' end of the siRNA guide dominate its affinity for target RNA¹⁹.

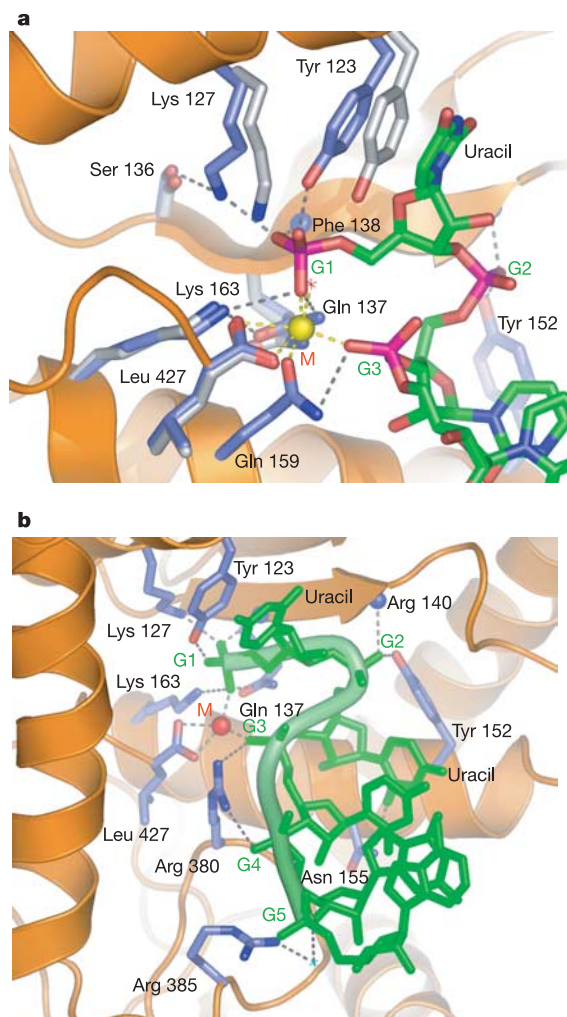


Figure 2 Interactions between AfPiwi and the guide RNA strand. **a**, Detailed view of the G1-binding pocket. Side chains coloured grey show conformations of selected residues in the absence of RNA¹². The concerted shift of Tyr 123 and Lys 127 suggests an induced-fit mechanism for phosphate binding. The divalent metal (M) is shown as a yellow sphere. **b**, View of interactions with the phosphate groups of G1 to G5. The target strand is omitted for clarity.

The G1-binding pocket and the putative slicer catalytic region^{11,12} lie at either end of the PIWI domain channel. Nucleotides G10–G16 of the guide and T10–T14 of the target are less clearly defined in our electron density maps. To visualize the trajectory of a full-length (19-nucleotide) siRNA guide bound to a mRNA target, we modelled the remaining nucleotides, assuming an A-form helix contiguous with the visible duplex (Fig. 3 and Supplementary Fig. 1). The weaker electron density for nucleotides G10 and G11 is compatible with this conformation (Fig. 1c). This model places the scissile phosphate of the target strand (linking T10 and T11)^{17,18} adjacent to the putative slicer catalytic region^{11,12} (Fig. 3 and Supplementary Fig. 3). The PIWI–RNA binding conformation is therefore consistent with our functional interpretation of the complex, being representative of a guide–target interaction, and provides insight into how the site-specificity of mRNA cleavage is

determined with reference to the 5' end of the guide^{17,18}.

Our structure might also reflect the initial phase of RISC association with a siRNA duplex; here the target strand would represent the passenger strand. The relative stabilities of the ends of an siRNA duplex determine the fate of the strands: the strand whose 5' end is located at the less stable end becomes the guide; the other (passenger) strand is discarded^{25,26}. Specific recognition of the siRNA-like 5' end, and partial unwinding of the duplex revealed from the AfPiwi–RNA complex, indicates that binding to the PIWI domain might be the first step in siRNA duplex unwinding in RISC; PIWI could therefore participate in strand selection. Consistent with this notion is evidence supporting a role for Argonaute 2 in siRNA unwinding in *Drosophila*²⁷. R2D2 has also been implicated in guide strand selection, binding with a preference to the more stable end of an siRNA duplex²⁸.

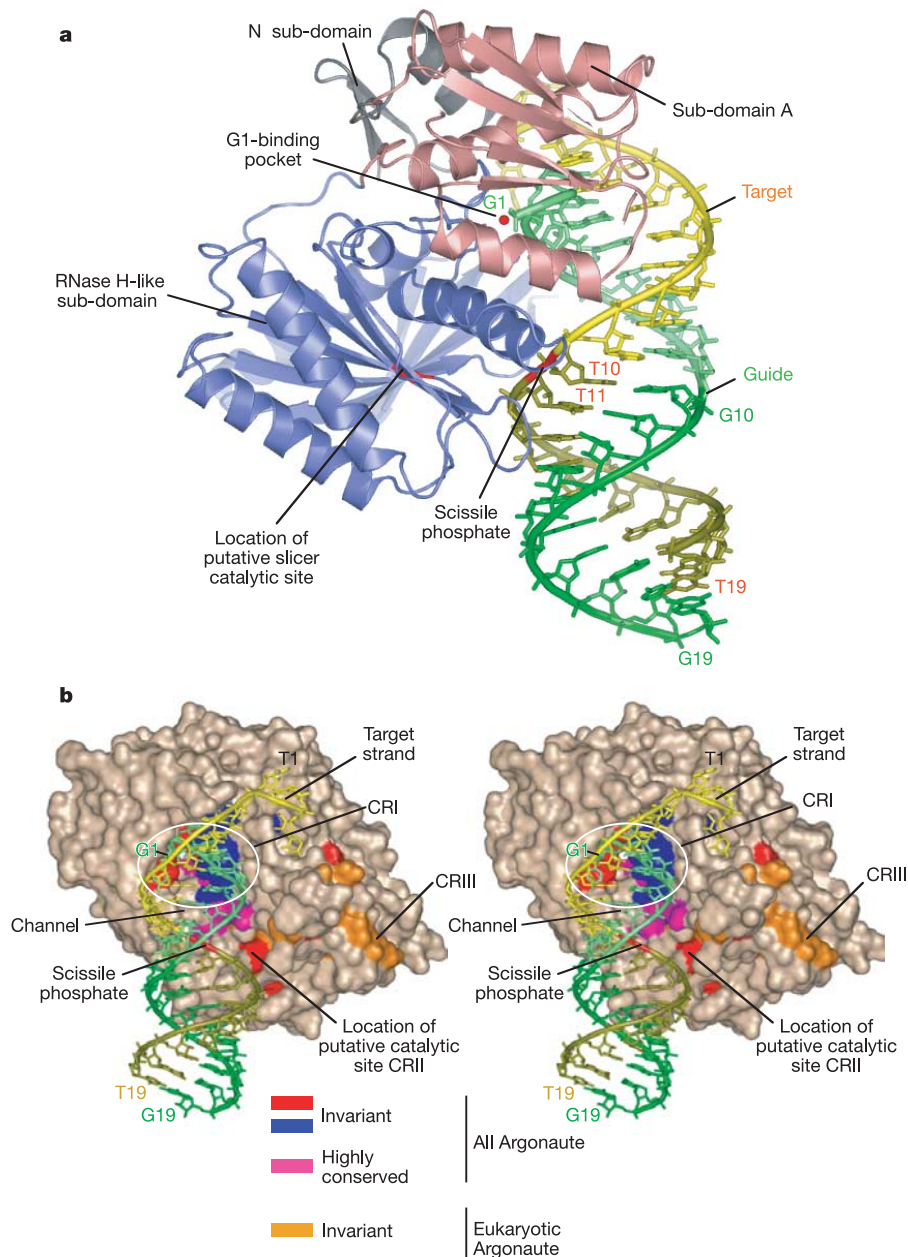


Figure 3 Proposed model for a 19-nucleotide guide–target duplex bound to AfPiwi. **a**, View showing the positioning of the scissile phosphate of the target strand (between T10 and T11, coloured red) with respect to the location of the putative slicer catalytic site. The modelled RNA region is coloured darker for both strands (G10 to G19 and T10 to

T19). PIWI sub-domains are shown¹². Catalytic Asp residues of the DDE motif^{11,12} are absent in AfPiwi. **b**, Stereo view with the molecular surface of AfPiwi colour-coded according to structural conservation (Supplementary Fig. 2). Conserved regions CRI, CRII and CRIII are indicated¹².

A role for the PIWI domain in mediating interactions with the 5' end of the guide RNA complements the function of the PAZ domain in recognizing RNA 3' overhangs^{7–10}. In *Pyrococcus furiosus* Argonaute, the PAZ domain is positioned over the channel linking the G1-binding pocket and the putative slicer site¹¹. The PAZ domain may function as a mobile unit that locks down over the guide–target duplex, providing additional constraints for slicing¹¹. However, when bound to a target mRNA, it is unlikely that the recessed guide 3' end would interact with the PAZ domain¹⁰; instead, we envisage the duplex emerging into solvent or the immediate environment of RISC.

Our structure rationalizes many molecular and biochemical data for mechanisms of RNA silencing, which, together with the universal conservation of the G1-binding pocket, indicate that the AfPiwi–RNA complex provides a representative picture of PIWI domain–RNA interactions. The view now emerges that Argonaute functions as a dynamic molecule mediating interactions with both the 5' and 3' ends of short guide RNA molecules, presenting them for recognition by cognate mRNA.

Note added in proof: In an accompanying paper³¹, Ma *et al.* use a different siRNA oligonucleotide to define the structure of an AfPiwi–RNA complex. Their structure is similar to ours, although less of the RNA molecule is defined. Consistent with our model for 5'-end guide recognition in eukaryotic Argonaute proteins, mutation of the G1-binding pocket in hAgo2 ablates slicer activity. □

Methods

Complex formation and crystallization

AfPiwi was purified as described¹², except that the protein was eluted at the final step in 5 mM HEPES pH 7.5, 800 mM NaCl, 2 mM MgCl₂, 1 mM MnCl₂, 8 mM dithiothreitol (DTT). The RNA oligonucleotide was purchased purified, deprotected and desalted from Pharmacia, and was resolubilized in 10 mM HEPES pH 7.5, 100 mM NaCl, 2 mM MgCl₂. Annealing was performed by incubation at 90 °C for 1 min followed by gradual cooling (1 °C per 5 min) from 65 °C to 4 °C. To form the complex, equal volumes of AfPiwi (260 µM) and annealed RNA (310 µM) were mixed and incubated at 22 °C for 30 min. Crystallization was performed with the hanging drop method at 14 °C. Crystal clusters were obtained overnight by mixing 1 µl of complex with 1 µl of 200 mM KCl, 10 mM CaCl₂, 50 mM sodium cacodylate pH 5.5, 15% PEG 4000, 5 mM DTT. Single crystals were obtained by streak-seeding into a similar drop containing 12% PEG 4000 after incubation overnight. For cryoprotection, crystals were passed through stabilization solutions containing increasing amounts of 2-methyl-2,4-pentane diol (to 17%) and then flash-frozen at 100 K. Data were collected at ID14.EH1, ESRF.

Structure determination

The structure (with two AfPiwi–RNA complexes per P1 unit cell) was solved by molecular replacement with PHASER²⁹ by using the native protein (PDB code 1W9H) as the search model¹². Some loops and all the RNA were rebuilt manually in COOT³⁰ and the complex was refined with REFMAC²⁹. A total of 508 water molecules were added in the final stages of refinement. The final *R* and *R*_{free} were 0.18 and 0.21, respectively (Supplementary Table 1). Terminal nucleotides of the target strands in the two RNA duplexes per unit cell adopt slightly different conformations (Supplementary Fig. 1). On the basis of crystallization conditions and refinement, we have assigned the metal ion within the G1-binding pocket as manganese.

Received 21 December 2004; accepted 17 February 2005; doi:10.1038/nature03462.

- Mello, C. C. & Conte, D. Jr Revealing the world of RNA interference. *Nature* **431**, 338–342 (2004).
- Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).
- Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* **16**, 2733–2742 (2002).
- Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
- Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
- Sontheimer, E. J. Assembly and function of RNA silencing complexes. *Nature Rev. Mol. Cell Biol.* **6**, 127–138 (2005).
- Yan, K. S. *et al.* Structure and conserved RNA binding of the PAZ domain. *Nature* **426**, 468–474 (2003).
- Song, J. J. *et al.* The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nature Struct. Biol.* **10**, 1026–1032 (2003).
- Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nature Struct. Mol. Biol.* **11**, 576–577 (2004).
- Ma, J. B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**, 318–322 (2004).

- Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**, 1434–1437 (2004).
- Parker, J. S., Roe, S. M. & Barford, D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J.* **23**, 4727–4737 (2004).
- Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).
- Nykanen, A., Haley, B. & Zamore, P. D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**, 309–321 (2001).
- Chiu, Y. L. & Rana, T. M. RNAi in human cells: basic structural and functional features of small interfering RNA. *Mol. Cell* **10**, 549–561 (2002).
- Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563–574 (2002).
- Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* **20**, 6877–6888 (2001).
- Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200 (2001).
- Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nature Struct. Mol. Biol.* **11**, 599–606 (2004).
- Hutvagner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056–2060 (2002).
- Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
- Stark, A., Brennecke, J., Russell, R. B. & Cohen, S. M. Identification of *Drosophila* MicroRNA targets. *PLoS Biol.* **1**, 397–409 (2003).
- Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
- Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are MicroRNA targets. *Cell* **120**, 15–20 (2005).
- Khvorov, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216 (2003).
- Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
- Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666 (2004).
- Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P. D. A protein sensor for siRNA asymmetry. *Science* **306**, 1377–1380 (2004).
- Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Ma, J.-B. *et al.* Structural basis for 5'-end-specific recognition of guide RNA by the A. *fulgidus* Piwi protein. *Nature* doi:10.1038/nature03514 (this issue).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to staff at ESRF for help with data collection. The work was funded by Cancer Research-UK, the Wellcome Trust and the Institute of Cancer Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.B. (david.barford@icr.ac.uk). The coordinates and structure factors have been deposited with the Protein Data Bank with accession numbers 2bgg and r2bggsf, respectively.

Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein

Jin-Biao Ma¹, Yu-Ren Yuan¹, Gunter Meister², Yi Pei², Thomas Tuschl² & Dinshaw J. Patel¹

¹Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

²Laboratory of RNA Molecular Biology, The Rockefeller University, New York, New York 10021, USA

RNA interference (RNAi) is a conserved sequence-specific gene regulatory mechanism^{1–3} mediated by the RNA-induced silencing complex (RISC), which is composed of a single-stranded guide RNA and an Argonaute protein. The PIWI domain, a highly conserved motif within Argonaute, has been shown to adopt an RNase H fold^{4,5} critical for the endonuclease cleavage activity of

RISC⁴⁻⁶. Here we report the crystal structure of *Archaeoglobus fulgidus* Piwi protein bound to double-stranded RNA, thereby identifying the binding pocket for guide-strand 5'-end recognition and providing insight into guide-strand-mediated messenger RNA target recognition. The phosphorylated 5' end of the guide RNA is anchored within a highly conserved basic pocket, supplemented by the carboxy-terminal carboxylate and a bound divalent cation. The first nucleotide from the 5' end of the guide RNA is unpaired and stacks over a conserved tyrosine residue, whereas successive nucleotides form a four-base-pair RNA duplex. Mutation of the corresponding amino acids that contact the 5' phosphate in human Ago2 resulted in attenuated mRNA cleavage activity. Our structure of the Piwi-RNA complex, and that determined elsewhere⁷, provide direct support for the 5' region of the guide RNA serving as a nucleation site for pairing with target mRNA and for a fixed distance separating the RISC-mediated mRNA cleavage site from the anchored 5' end of the guide RNA.

Small interfering RNAs (siRNAs), the cleavage products of RNase III enzyme Dicer, consist of RNA duplexes that contain two-nucleotide 3' overhangs and 5'-phosphorylated ends. These unique features at the termini of 19–23-base-pair duplexes, which distinguish siRNAs from other endogenous RNAs, are also required for their functional role in RNAi-mediated processes. Extensive structural and biochemical studies on PAZ (for PIWI/Argonaute/Zwille)-RNA recognition⁸⁻¹⁰, culminating in the structures of PAZ bound to single-stranded RNA¹¹ and siRNA-like duplexes¹², have established that the 3' overhang is specifically recognized by the PAZ domain, a module conserved in both Dicer and Argonaute proteins. Biochemical studies in *Drosophila melanogaster* lysates have shown that the 5' phosphate on the guide RNA is essential for its assembly into RISC^{13,14}, whereas related studies have established that the 5' phosphate of the partner strand is sensed in the process of RISC assembly by means of R2D2 (ref. 15). In addition, the cleavage-site position in the mRNA, paired to the guide RNA in the RISC

complex, is defined by the 5' end of the guide RNA strand¹⁶. Recent crystal structures have established that the PIWI domain, a highly conserved module in Argonaute proteins, adopts an RNase H fold⁴⁻⁶. The Piwi protein also contains within it a conserved basic pocket that probably constitutes the binding site for the 5' phosphate of the guide RNA⁵.

Here we describe the 2.5-Å crystal structure of the *A. fulgidus* Piwi protein (Fig. 1a) bound to a 5'-phosphate-containing 21-mer RNA, capable of self-complementary duplex formation (Fig. 1b). The crystallographic asymmetric unit contains two Piwi proteins, each bound to a short four-base-pair segment of duplex, positioned adjacent to the 5' phosphate and its attached unpaired nucleotide (Fig. 1c). The structure of the *A. fulgidus* Piwi protein, which consists of a short N-terminal element and two domains labelled A and B (Fig. 1a), is similar (root-mean-square deviation = 0.65 Å for Cα) in the free state⁵ and in the RNA-bound complex. The RNA, whose sequence from the 5' end is p-A1-G2-A3-C4-A5 (Fig. 1b), forms a short A-form duplex, involving pairing of the G2-A3-C4-A5 segment with its complement towards the 3' end of the sequence (Fig. 1c, d). The RNA duplex is positioned in a basic channel spanning the A-B inter-domain interface of the protein (Fig. 1e), whereas the 5' phosphate is inserted into a conserved basic pocket located primarily within domain A and the C terminus of domain B (Figs 1e and 2a). A network of hydrogen bonds involving the side chains of Y123, K127, Q137 and K163 and a bound divalent cation anchor the 5' phosphate into this pocket (Fig. 2a). The bound divalent cation, most probably Mg²⁺, is octahedrally coordinated to three protein-based oxygens, two nucleic-acid phosphate oxygens and a water molecule in the complex (Fig. 2b). This divalent cation, which is coordinated to the C-terminal carboxylate in an otherwise basic pocket, was also observed in the structure of the free *A. fulgidus* Piwi protein (identified as Cd²⁺)⁵, except that two coordinated chlorides are replaced by two phosphate oxygens on complex formation. These results are consistent with the formation of a preformed 5'-phosphate-binding pocket in Piwi, and by extension

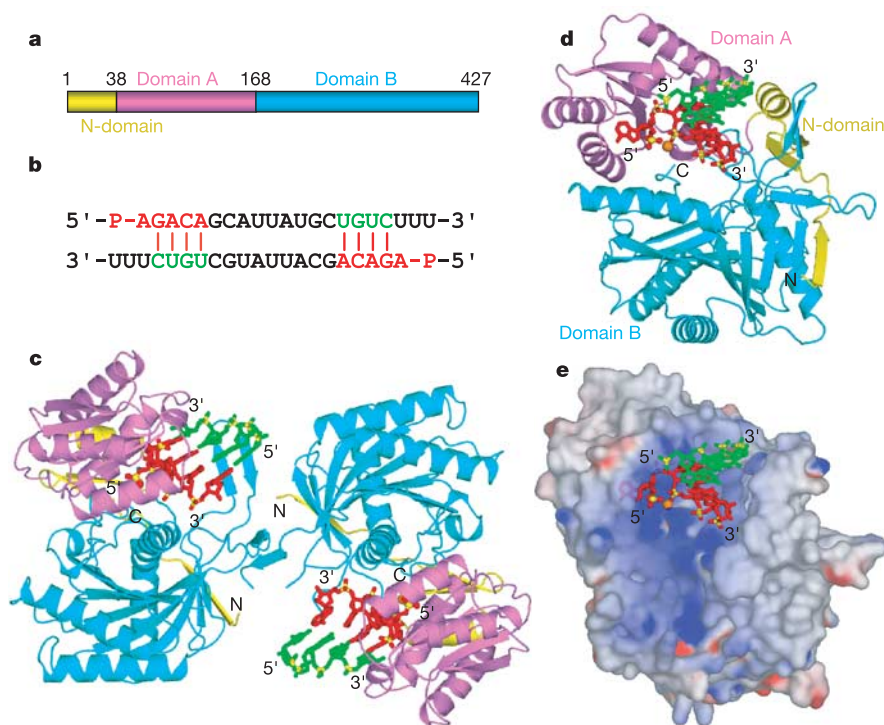


Figure 1 Crystal structure of the *A. fulgidus* Piwi-RNA complex. **a**, The *A. fulgidus* Piwi protein consists of a yellow N-domain (1–37), a magenta domain A (38–167) and a cyan domain B (168–427). **b**, The sequence and pairing alignment of the 21-mer RNA. The red and green segments are observed in the crystal structure of the complex, whereas

those in black are disordered. **c**, The relative alignments of the two Piwi proteins together with their bound RNAs within the asymmetric unit. **d**, **e**, Ribbon (**d**) and electrostatic (**e**) views of the complex. The RNA is shown in a stick representation.

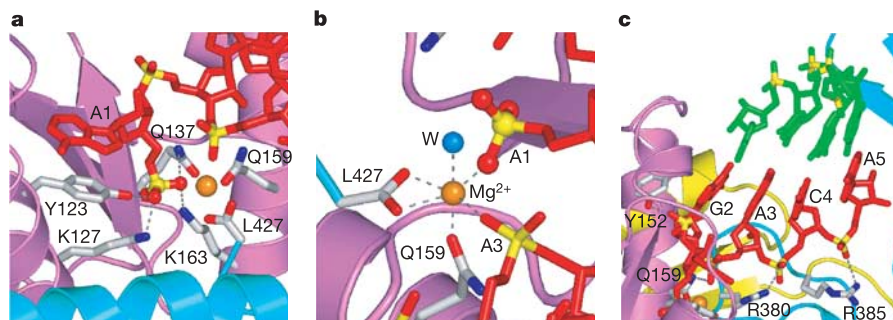


Figure 2 The 5'-phosphate-binding site in the *A. fulgidus* Piwi-RNA complex.

a, Positioning of the 5' phosphate in a basic pocket lined by residues of domain A (K127, Y123, Q137, Q159 and K163), domain B (C-terminal L427) and a bound divalent cation in orange. Bases A1 and G2 are displayed apart, with unpaired A1 stacked on Y123. **b**, The octahedral coordination of the bound divalent cation, presumably Mg^{2+} , involving protein

and RNA oxygens and a water molecule. **c**, Positioning of the short four-base-pair duplex segment involving the G2-A3-C4-A5 segment (red) with its complementary partner (green). Note intermolecular hydrogen bonds to the non-bridging phosphate oxygens of the A1-G2-A3-C4-A5 segment.

in Argonaute proteins. They also support an earlier prediction of the potential location of the 5'-phosphate-binding pocket within the Piwi scaffold⁵.

We propose that it is the guide RNA strand whose 5' phosphate is anchored in the basic pocket of the Piwi protein. The A1 base at the 5' end of the guide RNA strand is unpaired and stacks on the aromatic ring of invariant Y123 (Fig. 2a), thereby accounting for non-sequence-specific recognition of the first nucleotide within the 5'-end-binding pocket. The non-bridging oxygens of the backbone phosphates at the A1-G2, G2-A3, A3-C4 and C4-A5 steps are hydrogen-bonded to the side chains of Y152, Q159, R380 and R385,

respectively (Fig. 2c). The two paired RNA strands bind in a strongly asymmetric manner, such that the 5' end of the guide RNA strand (coloured red in Fig. 2c) interacts extensively with the protein, burying $1,500 \text{ \AA}^2$ of protein surface, whereas the mRNA strand (coloured green in Fig. 2c) positioned opposite it has minimal contacts with the protein, burying $350 \text{ \AA}^2$ of protein surface.

We have generated a set of single and double mutants of *A. fulgidus* Piwi protein residues that contact the 5' phosphate and attached A1 residue of the bound RNA, in an effort to investigate the importance of these highly conserved residues to RNA recognition. Filter binding studies (averages of three runs)

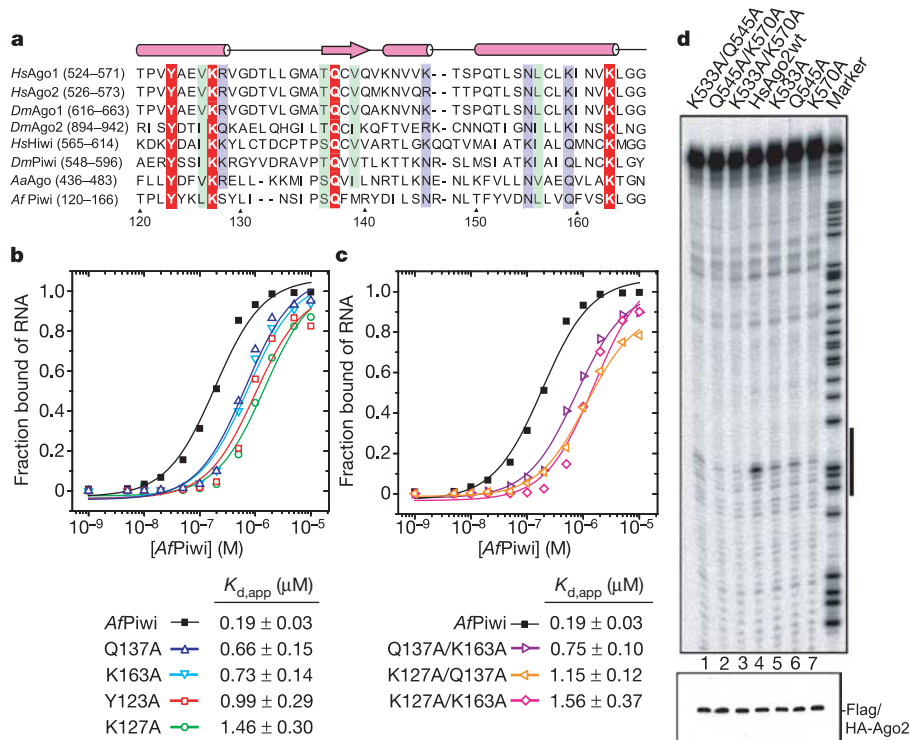


Figure 3 Mutation studies of conserved 5'-phosphate-binding residues in *A. fulgidus* Piwi protein and human Ago2 protein. **a**, Sequence alignment of conserved 5'-phosphate-binding residues across various species of Piwi and Argonaute proteins. Prefix Hs, human; Dm, *Drosophila*; Aa, *Aquifex aeolicus*; Af, *Archaeoglobus fulgidus*. **b**, Representative binding curves (average of three runs) for *A. fulgidus* Piwi single-site mutants with 21-nucleotide RNA labelled with 5'- ^{32}P by double-filter binding assay. The deduced

$K_{d,app}$ values are listed below. **c**, The corresponding binding curves for *A. fulgidus* Piwi dual-site mutants. **d**, The cleavage activities of wild-type human Ago2 and its mutants. The black bar at the right of the image represents the region of the target RNA complementary to the siRNA used. The expression levels of the proteins used in the assays were assessed by western blotting with anti-HA antibody and are shown in the lower panel.

establish that the apparent binding affinity for RNA ($K_{d,app}$) decreases from threefold to sevenfold on proceeding from the wild-type protein (0.19 μ M) to mutants Q137A (0.66 μ M), K163A (0.73 μ M) and K127A (1.46 μ M), all of which should disrupt hydrogen bonding to the 5' phosphate, and mutant Y123A (0.99 μ M), which should disrupt stacking with the A1 residue (Fig. 3b). The same trend in binding affinity loss is observed for double mutants Q137A/K163A (0.75 μ M), K127A/Q137A (1.15 μ M) and K127A/K163A (1.56 μ M) (Fig. 3c). These results indicate that 5' phosphate stabilization might involve multiple hydrogen-bonding interactions and that the disruption of any single interaction or pair of such interactions results in only a modest decrease in binding affinity.

Because the *A. fulgidus* Piwi protein shares significant sequence homology with human Argonaute proteins (Fig. 3a), we mapped the 5'-end-binding pocket in human Ago2, the RISC component that mediates small RNA-guided cleavage of target mRNA^{6,17}, and set out to investigate whether mutation of key residues lining the 5'-end-binding pocket would affect human Ago2 mRNA cleavage activity. Guided by the crystal structure of the *A. fulgidus* Piwi protein–RNA complex, we introduced single as well as double alanine mutations at the corresponding residues located in the putative 5'-end-binding pocket of human Ago2, namely K533, Q545 and K570 (Fig. 3a). The wild-type human Ago2 guided the cleavage of the radiolabelled target RNA efficiently (Fig. 3d, lane 4). The single alanine mutants, K533A, Q545A and K570A, showed reduced cleavage activity (Fig. 3d, lanes 5 to 7), whereas the double alanine mutants manifested more severe defects in cleavage capacity (Fig. 3d, lanes 1 to 3). These results indicate that the 5'-end-binding pocket observed in *A. fulgidus* Piwi protein is conserved in human Ago2 and is important for its small RNA guided cleavage activity.

We have also directly investigated the effect of 5'-phosphorylation of RNAs on binding affinity. These results, based on a double filter-binding competition assay, establish that the 5'-phosphate-containing RNA (Supplementary Fig. S2a, panels labelled B) binds the *A. fulgidus* Piwi protein about an order of magnitude more tightly than its non-phosphorylated counterpart (Fig. S2a, panels labelled A). It therefore seems that 5'-phosphorylation indeed provides additional affinity for guide RNA recognition of *A. fulgidus* Piwi protein.

There are no specific intermolecular contacts directed to the 2'-OH groups of the RNA in the crystal structure of the *A. fulgidus* Piwi protein–RNA complex, indicating that the Piwi protein might also be targeted by other types of nucleic acid. We have investigated the apparent binding affinities of *A. fulgidus* Piwi protein to both single-stranded and double-stranded 21-mer DNA and RNA oligonucleotides with the use of double filter-binding assays (Supplementary Fig. S2b). The binding is 20-fold tighter for single-stranded DNA than for ssRNA, with the binding affinity decreasing on proceeding from DNA–DNA to RNA–DNA to RNA–RNA duplexes. The preference for ssDNA and dsDNA of the archaeal Piwi protein was unexpected and will need further investigation. Nevertheless, our results are consistent with previous observations that the 2'-OH is not required for RNAi-mediated events¹⁸. It is also conceivable that other domains within Argonaute proteins and/or other RISC components could provide additional interactions that discriminate between nucleic acid types.

Our structure revealed that the stacked bases 2 to 5 at the 5' end of the guide RNA are accessible for pairing with their complementary counterparts on the mRNA. Bioinformatics and biochemical analysis of microRNA (miRNA) genes have shown a high degree of sequence conservation restricted to positions 2–8, indicative of a significant contribution of the 5' end to specificity and activity^{19–21}. Our structure supports this concept, in which the mRNA would initially nucleate with the 5' end of the protein-bound guide RNA strand^{19,22}, followed by zippering up of the remaining segment to form the guide RNA–mRNA duplex. The observation that the first

base on the guide RNA strand is not available for pairing with mRNA in the crystal structure of the complex is consistent with earlier biochemical and kinetic data showing that disruption of this pair has no effect on cleavage activity and under some conditions even favours target cleavage²².

More than half of the bound 21-mer is disordered in the crystal structure of the *A. fulgidus* Piwi protein–RNA complex (Fig. 1c). We have therefore constructed a model in which the trajectory of the four-base-pair duplex observed in the crystal structure of the complex was extended to form two turns of A-form helix (Fig. 4a). This duplex ends up being positioned within a basic channel⁵ that spans one face of both domains A and B of a monomeric Piwi protein (Fig. 4b). The proposed mRNA cleavage site is positioned opposite the putative catalytic residues associated with the RNase H fold of domain B of the Piwi protein^{4–6} (Fig. 4a), although the archaeal *A. fulgidus* Piwi protein lacks the DD components of the conserved DDE motif, normally associated with RNase H cleavage activity²³ (Supplementary Fig. S1).

Previous studies have established that the RISC complex cleaves the mRNA strand in a site-specific manner between positions 10 and 11 as defined from the 5' end of the guide strand¹⁶. This implies that the catalytic cleavage site is measured from the bound 5' end of the guide strand by a fixed distance, a feature consistent with the anchoring of both the 5' phosphate and its adjacent nucleotide of the guide strand, as observed in our structure of the *A. fulgidus*

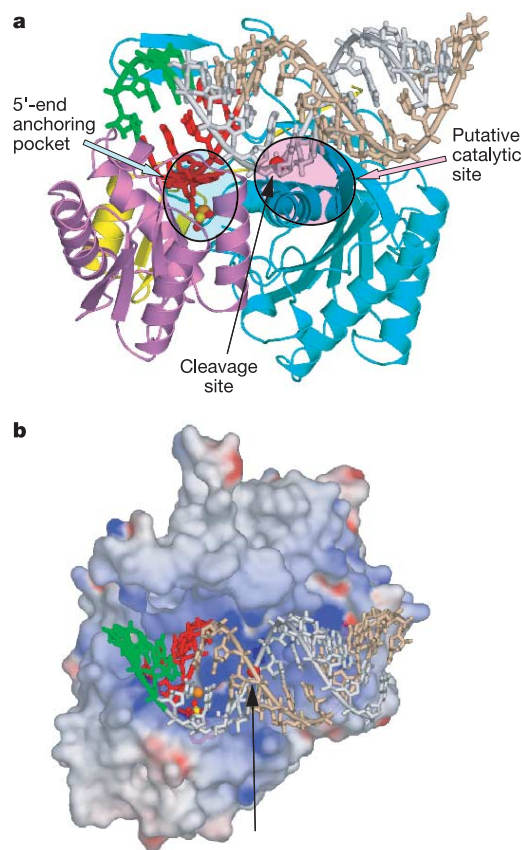


Figure 4 Model of *A. fulgidus* Piwi protein (AfPiwi) bound to an 18-base-pair RNA duplex with 5'-phosphorylated single-nucleotide overhangs. **a**, The 5'-end nucleotide and four-base-pair duplex (coloured in red and green for guide strand and target strand, respectively) observed in the crystal structure of the complex were extended by 14-base-pair A-form duplex (coloured in tan and white for guide strand and target strand, respectively). The 5' end anchoring pocket and putative catalytic site are circled, and the phosphate cleavage site is marked by a red ball. **b**, Electrostatic surface view of the model of the AfPiwi–dsRNA complex.

Piwi protein–RNA complex (Figs 2a and 4a).

Argonautes, which contain conserved PAZ and PIWI domains, seem to be the sole proteins required for RISC-mediated activities²⁴. Active RISCs, which contain only a single guide RNA strand²⁵, are capable of directing multiple rounds of site-specific target mRNA cleavage²⁶. Both our structure of the *A. fulgidus* Piwi protein–RNA complex and that determined by others⁷ provides the first structural insights into 5′ end recognition of the guide RNA strand, the nucleation step associated with guide-RNA–mRNA recognition and the distance-based identification of the mRNA cleavage site. Further attempts at structural characterization of full-length Argonaute–RNA complexes, if successful, could provide structure-based information related to the catalytic cleavage mechanism, a central event of RNA interference. □

Methods

Protein and RNA preparations

The full-length *A. fulgidus* Piwi cDNA was amplified by polymerase chain reaction (PCR) from *Archeoglobus fulgidus* genomic DNA (ATCC) and cloned into pET28a (Novagen) with an amino-terminal His₆ tag. The protein was overexpressed in the *Escherichia coli* host cell BL21–Gold(DE3) (Stratagene) induced by 1 mM isopropyl β-D-thiogalactoside at 20 °C, and purified by Ni-chelating and heparin Hitrap columns. After gel-filtration chromatography on Superdex 200 (Amersham Biosciences), the pure *A. fulgidus* Piwi with attached His tag was concentrated and stored at 30 mg ml^{−1} in 10 mM Tris–HCl pH 8.0, 1 M NaCl and 1 mM dithiothreitol. Selenomethionine-labelled protein was produced by inhibiting the methionine biosynthesis pathway in the host cell. Site-directed mutations were made by a two-step PCR-based overlap extension method and confirmed by DNA sequencing. All the mutants were purified as described for the wild-type protein except that the Heparin column step was omitted. RNAs were synthesized chemically in our laboratory or purchased from Dharmacon, then deprotected and purified by denaturing polyacrylamide-gel electrophoresis.

Crystallization and data collection

Crystals of native and selenomethionine-substituted *A. fulgidus* Piwi protein–21-mer RNA complex were grown by hanging-drop vapour diffusion at 20 °C. About 15 mg ml^{−1} protein–RNA complex in 0.1 M KCl, 10 mM HEPES–K pH 7.5 buffer was mixed 1:1 with a well solution containing 6% polyethylene glycol 4000, 40 mM KCl, 10 mM MgCl₂ in 50 mM sodium cacodylate pH 6.5. These crystals grew to a maximum size of 0.2 mm × 0.2 mm × 0.2 mm over the course of 10 days.

For data collection, crystals were flash frozen (100 K) in the above reservoir solution supplemented with 30% glycerol with a 10% step increase. A total of 720 frames of 0.5° oscillation were collected on the 2.5 Å crystals of selenomethionine-labelled *A. fulgidus* Piwi protein–21-mer RNA complex at 0.9776 Å at National Synchrotron Light Source (NSLS) beamline X26C, Brookhaven National Laboratory. The data were processed with HKL2000 (<http://www.hkl-xray.com/>). The crystal belonged to space group R3, with unit cell parameters listed in Supplementary Table S1. There were two complex molecules per asymmetric unit, with about 55% solvent content.

Structure determination

The structure of the *A. fulgidus* Piwi protein–21-mer RNA complex was solved by using the free *A. fulgidus* Piwi protein structure (PDBID: 1W9H) as the search model and Molrep (CCP4 package)²⁷ was used to locate the two copies of the *A. fulgidus* Piwi protein. The extra density corresponding to the 5′ terminus of the RNA and several additional base pairs could be clearly traced without ambiguities from the earliest stage. The *A. fulgidus* Piwi protein–RNA complex model was rebuilt with the program O²⁸ and refined with CNS (<http://cns.csb.yale.edu/v1.1/>) and REFMAC (CCP4 package)²⁷. The R-free set contained 5% of the reflections chosen at random. Refinement statistics are given in Supplementary Table S1. The protein comprises residues 2–427 and the RNA comprises the 5′ phosphate, the A1 residue and four additional base pairs involving residues at the 5′ end. Disordered regions within the *A. fulgidus* Piwi protein included loop segments 304–308 and 331–341 for molecule A and 332–338 for molecule B in the asymmetric unit. The middle part of the duplex RNA (nine base pairs) and three nucleotides at the 3′ end cannot be traced in our final density map. The figures were prepared with PyMOL (<http://pymol.sourceforge.net/>).

Double filter-binding assays

The affinity of the *A. fulgidus* Piwi and its mutants for various oligonucleotides was detected and quantified with a double filter-binding assay²⁹. The reactions containing 5′-³²P-labelled nucleic acids and proteins were applied to filters containing two membranes: a protein-binding Protran BA-85 nitrocellulose membrane and a nucleic-acid-binding Hybond-N⁺ nylon membrane. After air-drying, the filters were quantified on a PhosphorImager. The data were analysed as reported previously³⁰; details of experimental protocols are listed in Supplementary Methods.

Cleavage assays using human Ago2 and its mutants

The Flag/haemagglutinin (HA)-tagged wild-type and mutant constructs were transiently transfected into HEK-293 cells, and Ago2 proteins were immunoprecipitated from the cell lysates with anti-Flag antibody. The expression levels of the wild-type and mutant Ago2 proteins were assessed by western blotting with anti-HA antibody (Fig. 3c, lower panel).

The stringently washed beads carrying equal amounts of bound Flag/HA-tagged human Ago2 proteins were preincubated with a single-stranded siRNA to reconstitute RISC activity. After an additional washing step to remove the unbound siRNA, the beads were incubated with a ³²P-cap-labelled target RNA. After incubation for 1 h the RNA was recovered from the beads and analysed by denaturing gel electrophoresis. Details of the experimental protocols are listed in Supplementary Methods.

Received 6 February; accepted 10 March 2005; doi:10.1038/nature03514.

- Hutvagner, G. & Zamore, P. D. RNAi: nature abhors a double-strand. *Curr. Opin. Genet. Dev.* **12**, 225–232 (2002).
- Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
- Meister, G. & Tuschl, T. Mechanism of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).
- Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**, 1434–1437 (2004).
- Parker, J. S., Roe, S. M. & Barford, D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J.* **23**, 4727–4737 (2004).
- Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).
- Parker, J. S., Roe, S. M. & Barford, D. Structural insights into mRNA recognition from a PIWI-domain–siRNA guide complex. *Nature* doi:10.1038/nature03462 (this issue).
- Yan, K. S. *et al.* Structure and conserved RNA binding of the PAZ domain. *Nature* **426**, 468–474 (2003).
- Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain. *Nature* **426**, 465–469 (2003).
- Song, J. J. *et al.* The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nature Struct. Biol.* **10**, 1026–1032 (2003).
- Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain. *Nature Struct. Mol. Biol.* **11**, 576–577 (2004).
- Ma, J. B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**, 318–322 (2004).
- Nykänen, A., Haley, B. & Zamore, P. D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**, 309–321 (2001).
- Tomari, Y. *et al.* RISC assembly defects in the *Drosophila* RNAi mutant Armitage. *Cell* **116**, 831–841 (2004).
- Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P. D. A protein sensor for siRNA asymmetry. *Science* **306**, 1377–1380 (2004).
- Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200 (2001).
- Meister, G. *et al.* Human Argonaute2 mediates RNA cleavage targeted by miRNA and siRNAs. *Mol. Cell* **15**, 185–197 (2004).
- Chiu, Y. L. & Rana, T. M. siRNA function in RNAi: a chemical modification analysis. *RNA* **9**, 1034–1048 (2003).
- Mallory, A. C. *et al.* MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5′ region. *EMBO J.* **23**, 3356–3364 (2004).
- Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
- Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
- Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nature Struct. Mol. Biol.* **11**, 599–606 (2004).
- Yang, W. & Steitz, T. A. Recombining the structures of HIV integrase, RuvC and RNase H. *Structure* **3**, 131–134 (1995).
- Rand, T. A., Ginalski, K., Grishin, N. V. & Wang, X. Biochemical identification of Argonaute2 as the sole protein required for RNA-induced silencing complex activity. *Proc. Natl Acad. Sci. USA* **101**, 14385–14389 (2004).
- Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563–574 (2002).
- Hutvagner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056–2060 (2002).
- Collaborative Computational Project no. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, G. J. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Wong, I. & Lohman, T. M. A double-filter method for nitrocellulose-filter binding: Application to protein–nucleic acid interactions. *Proc. Natl Acad. Sci. USA* **90**, 5428–5432 (1993).
- Sagar, M. B., Lucast, L. & Doudna, J. A. Conserved but nonessential interaction of SRP RNA with translation factor EF-G. *RNA* **10**, 772–778 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Saxena and personnel at synchrotron beamline X26C of the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory, for their assistance. Use of the NSLS beamline is supported by the US Department of Energy, Basic Energy Sciences, Office of Science. D.J.P. is supported by funds from the Abby Rockefeller Mauze Trust and the Dewitt Wallace and Maloris Foundations, and T.T. is supported by a National Institutes of Health grant.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.P. (pateld@msskcc.org). Coordinates have been deposited in the Protein Data Bank under accession code 1YTU.

Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB

Daniel A. Koster¹, Vincent Croquette², Cees Dekker¹, Stewart Shuman³ & Nynke H. Dekker¹

¹Kavli Institute of Nanoscience, Faculty of Applied Sciences, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

²Laboratoire de Physique Statistique, Ecole Normale Supérieure, 75005 Paris, France

³Molecular Biology Program, Sloan-Kettering Institute, New York, New York 10021, USA

Topoisomerases relieve the torsional strain in DNA that is built up during replication and transcription. They are vital for cell proliferation^{1–3} and are a target for poisoning by anti-cancer drugs^{4,5}. Type IB topoisomerase (TopIB) forms a protein clamp around the DNA duplex^{6–8} and creates a transient nick that permits removal of supercoils. Using real-time single-molecule observation, we show that TopIB releases supercoils by a swivel mechanism that involves friction between the rotating DNA and the enzyme cavity: that is, the DNA does not freely rotate. Unlike a nicking enzyme, TopIB does not release all the supercoils at once, but it typically does so in multiple steps. The number of supercoils removed per step follows an exponential distribution. The enzyme is found to be torque-sensitive, as the mean number of supercoils per step increases with the torque stored in the DNA. We propose a model for topoisomerization in which the torque drives the DNA rotation over a rugged periodic energy landscape in which the topoisomerase has a small but quantifiable probability to religate the DNA once per turn.

Type IB topoisomerases alter DNA topology by cleaving and rejoining one strand of the DNA duplex¹, and are able to remove both positive and negative supercoils. *In vivo*, TopIB removes positive supercoils generated in advance of the replication fork⁹. Cleavage occurs via a transesterification reaction in which the scissile phosphodiester is attacked by a tyrosine of the enzyme, resulting in the formation of a DNA–(3′-phosphotyrosyl)–enzyme intermediate and the expulsion of a 5′-OH DNA strand. In the rejoining step, the DNA 5′-OH group attacks the covalent intermediate, resulting in expulsion of the active-site tyrosine and restoration of the DNA phosphodiester backbone.

On the basis of structural¹⁰ and kinetic¹¹ data, a mechanism has been proposed for TopIB, whereby supercoils are relaxed by swivelling of the DNA about the phosphodiester opposite the nick (Fig. 1a–c). The clamp-like structure of human TopIB around the DNA duplex suggests that the DNA cannot swivel unhindered, and consequently a ‘controlled rotation’ model has been proposed^{10,11}. The swivel mechanism of TopIB action is in stark contrast to the protein-assisted strand-passage mechanism of type IA and type II topoisomerases, in which there is an obligate step-size of 1 or 2 supercoils removed per cleavage–religation cycle, respectively^{12–18}.

Our experimental strategy entails anchoring a single, continuous linear double-stranded (ds)DNA molecule between a glass surface and a paramagnetic bead. We use a pair of magnets to apply a stretching force F and a degree of supercoiling σ by translating a pair of magnets in the vertical direction or rotating them about their axis, respectively¹⁹. The height of the bead from the surface is equal to the extension of the DNA molecule. Figure 1d plots the DNA extension as a function of force and torque. In a typical measurement, the DNA molecule is prepared in a positively supercoiled state ($\sigma > 0$), which corresponds to a short extension.

When vaccinia TopIB, a prototypical eukaryotic type IB topoisomerase, is added, we observe discrete, step-wise increases in the extension (Fig. 2a). Each step signifies the removal of DNA supercoils during a single cleavage–religation cycle by a single TopIB enzyme. Using the rotation curve (Fig. 1d), we convert the changes

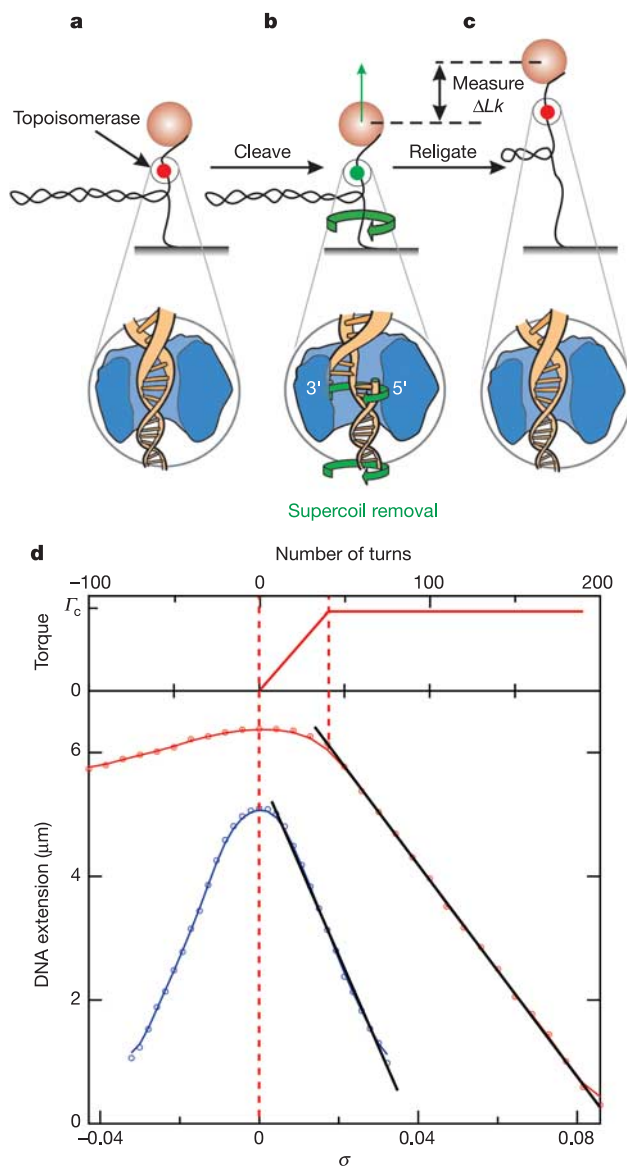


Figure 1 Single-molecule assay for measuring DNA supercoil removal by topoisomerase.

a, Type IB topoisomerase is bound non-covalently to a supercoiled DNA molecule. **b**, Topoisomerase establishes a covalent bond to the DNA, creating a nick that allows for rotation of the DNA about the remaining, intact DNA strand (green arrow). In this way, supercoils are removed. **c**, The DNA has been religated by the topoisomerase, terminating the release of supercoils. The height of the bead above the surface has increased, proportional to the number of supercoils removed (ΔLk). **d**, The behaviour of dsDNA under torsion is dependent on the stretching force (data shown for half-length bacteriophage λ DNA). Increasing σ at 1 pN (red curve), the extension initially remains constant and the torque builds up linearly with σ (Twist regime). After a critical buckling torque T_c , the torque saturates and the DNA forms plectonemic supercoils. It then contracts linearly with σ (Writhe regime), with a slope of 37 nm per turn (1 pN) and 65 nm per turn (0.2 pN, blue curve). At low stretching force (0.2 pN) the DNA extension is decreased irrespective of the rotation direction as the molecule is supercoiled. At a higher force (1 pN), this occurs only for positive rotations ($\sigma > 0$); at negative rotations, the DNA denatures. Solid coloured lines in rotation curves are splines through experimental data points, black lines are fits to a linear function.

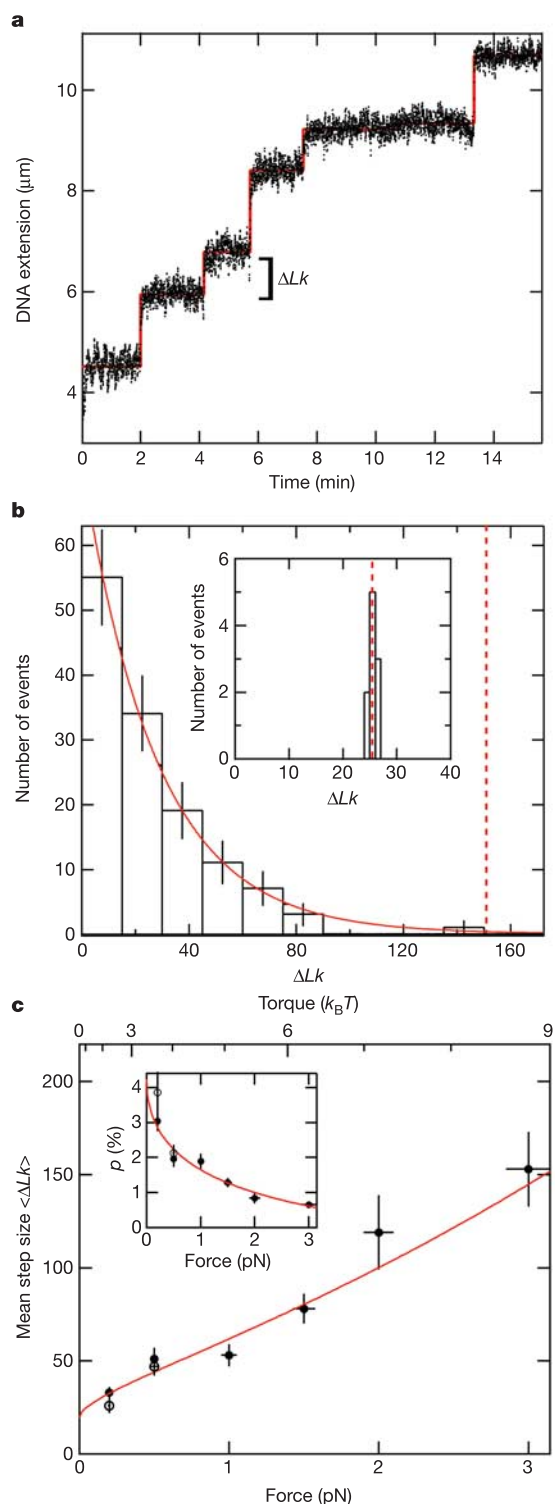


Figure 2 Real-time enzymatic activity and step-size distribution for TopIB acting on a single DNA molecule. **a**, Each time TopIB removes supercoils from the DNA, a step is observed in the DNA extension. **b**, $P(\Delta Lk)$ for TopIB and for nicking enzyme (inset), in units of ΔLk . Vertical dashed red lines show the number of pleconemic supercoils initially incorporated into the DNA. The solid red line is a fit of equation (1) to the data. Error bars denote the square root of the number of events. **c**, $\langle \Delta Lk \rangle$ as a function of applied force and torque. The solid line shows the fit of equation (2) to the data. The inset shows the force-dependence of p (see text). Solid and open circles represent data taken at positive and negative supercoils, respectively.

in DNA extension to a number of rotations removed from the DNA, which equals the change in the linking number ΔLk . The distribution of ΔLk , $P(\Delta Lk)$, has a mean far greater than unity (Fig. 2b) and is well fitted by an exponential (red line in Fig. 2b):

$$P(\Delta Lk) \approx \exp(-\Delta Lk / \langle \Delta Lk \rangle) \quad (1)$$

where $\langle \Delta Lk \rangle$ refers to the mean change in linking number. At higher stretching forces, the functional forms of the distributions are unchanged, but the corresponding values for $\langle \Delta Lk \rangle$ increase significantly (Fig. 2c, solid circles). The functional forms of the distributions are also found to be insensitive to the sign of the supercoiling (data not shown). Within the experimentally accessible regime at negative supercoiling (see Fig. 1d), the values of $\langle \Delta Lk \rangle$ for negative supercoils (Fig. 2c, open circles) are similar to those measured for positive supercoils. In contrast to vaccinia TopIB, a control measurement of $P(\Delta Lk)$ for a nicking enzyme shows a peak that coincides with the number of supercoils initially applied to the DNA molecule (Fig. 2b, inset). Unlike vaccinia TopIB, the nicking enzyme does not possess the ability to religate DNA and therefore all supercoils are released at once.

The exponential functional form of $P(\Delta Lk)$ for vaccinia TopIB can be understood by considering the DNA as it rotates during supercoil release. After nicking of the DNA by transesterification of vaccinia TopIB to the DNA 3'-phosphate end, the DNA rotates inside the enzyme cavity and its 5'-OH end passes the tyrosine-3'-DNA adduct once every turn. At each pass, there is a finite probability p for vaccinia TopIB to religate the DNA and a probability $q \equiv 1 - p$ that the enzyme does not religate the DNA. In accordance with this, the probability distribution for observing steps of size ΔLk is given by the discrete probability function of the geometric distribution²⁰ ($\Delta Lk = 1, 2, \dots$): $P(\Delta Lk) = pq^{\Delta Lk-1} = p(1-p)^{\Delta Lk-1}$, where $\langle \Delta Lk \rangle \equiv 1/p$. The continuum limit of this distribution is equation (1). We find that by increasing the stretch-

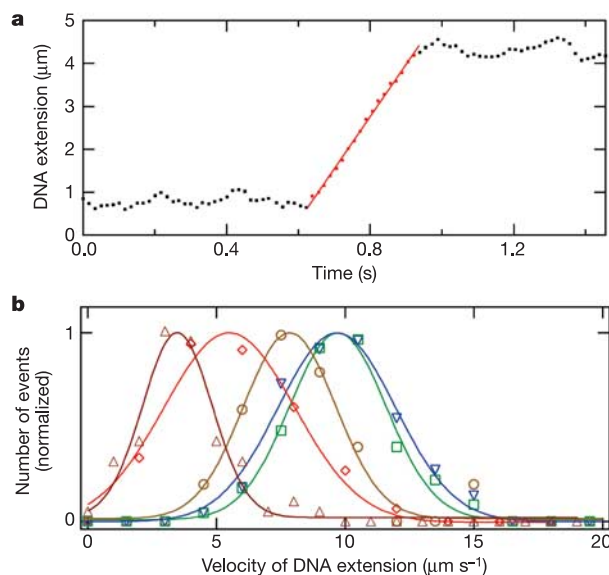


Figure 3 Measurement of velocity of DNA extension during supercoil release. **a**, A step in DNA extension as a consequence of supercoil removal is characterized by a linear increase in DNA extension (red points) and is fitted with a linear function to obtain the rate (solid red line). **b**, Velocity of DNA extension distributions taken at 0.2 pN for human TopIB (dark red triangles, $\langle v \rangle = 4.1 \pm 0.2 \mu\text{m s}^{-1}$), wild-type TopIB (red diamonds, $\langle v \rangle = 6.7 \pm 0.2 \mu\text{m s}^{-1}$), the TopIB Y70A mutant (beige circles, $\langle v \rangle = 8.9 \pm 0.6 \mu\text{m s}^{-1}$), nicking enzyme (blue triangles, $\langle v \rangle = 10.5 \pm 0.2 \mu\text{m s}^{-1}$) and for a mix of TopIB mutant Y274F and nicking enzyme (green squares, $\langle v \rangle = 10.5 \pm 0.2 \mu\text{m s}^{-1}$). Solid lines are gaussian fits to the data. Means are numerical averages with the corresponding s.e.m.

ing force from 0.2 pN to 3 pN, the probability of religation per turn decreases from 3% to 0.7% (Fig. 2c, inset). Because a strand-passage mechanism implies that $\langle \Delta Lk \rangle$ is independent of the stretching force¹⁷, we conclude that our data are inconsistent with such a model. Our results are, however, fully consistent with a swivel mechanism.

We are able to resolve in real-time the velocity of DNA extension during supercoil removal. We measure this both for vaccinia TopIB and for endonucleolytic cleavage of one strand by nicking enzymes (Fig. 3). The velocity value is a measure of the rate at which the DNA swivels in the topoisomerase cavity as the supercoils are released. The data in Fig. 3b clearly show that vaccinia TopIB (red diamonds) slows down the DNA rotation rate compared to the unhindered rate observed for the nicking enzymes (blue triangles). Control experiments with two different vaccinia TopIB mutants and with human TopIB provide further evidence that the decreased rotation rate in the TopIB reaction is caused by friction between the topoisomerase and the rotating DNA. First, we excluded the possibility that multiple topoisomerases bound to the DNA at sites other than the cleavage site were interacting with each other²¹ in a manner that caused a decreased rotation rate. This was accomplished by a mixing experiment in which the nicking enzyme was reacted with the DNA in the presence of the vaccinia TopIB active-site mutant Y274F (green squares), which binds DNA noncovalently but is incapable of transesterification²². The rotation rate observed in the mixed reaction was indistinguishable from the rate observed with the nicking enzyme only. Second, we measured the effect of the vaccinia Y70A mutation on the DNA rotation rate. Tyr 70 is located on the concave surface of the amino-terminal domain of vaccinia TopIB, which clamps over the DNA duplex in the major groove on the helical face opposite the scissile phosphodiester²³. Owing to the loss of this amino-acid side chain lining the protein–DNA interface, we observe an increase in the DNA rotation rate (beige circles) compared with wild-type vaccinia TopIB. Extending these measurements to human TopIB (dark red triangles), we find that the DNA rotation rate is lower than for wild-type vaccinia TopIB. Human TopIB encircles the DNA duplex fully⁷, but vaccinia TopIB has the form of a C-shaped clamp^{6,8}. Accordingly, human TopIB presumably offers less freedom for the DNA to rotate than vaccinia TopIB. These data provide what is, to our knowledge, the first direct

measurement of friction^{24,25} in the TopIB relaxation mechanism.

We propose a model that describes the effect of friction and torque on enzymatic activity. Our model has three ingredients. First, the rotation of the DNA inside the enzyme clamp is not free but is hindered by friction, as indicated by our velocity measurements. This is modelled by a random walk over a rugged free-energy landscape (Fig. 4), with the rotation angle θ between the 5'-OH end of the noncovalently held strand and the tyrosine-3'-DNA adduct as the reaction coordinate. The DNA rotation is not smoothly continuous, but the free energy profile and accordingly the rotation rate vary during a single rotation. This variability in rotation speed could stem from the notion that the cross-sectional size of the DNA at the nick changes dramatically, from 2 nm at $\theta = 0$ to about 4 nm at $\theta = 180^\circ$. We make no assumptions about the exact shape of the energy landscape. The rate k across each of the barriers with height ΔG in this landscape follows an Arrhenius relation, $k \approx \exp(-\Delta G/k_B T)$, where k_B is the Boltzmann constant and T the temperature in Kelvin.

Second, the mechanically applied constant torque Γ_c drives the uncoiling. It is modelled by tilting the entire energy landscape by $-\Gamma_c \theta$. This decreases ΔG by an amount $\Gamma_c \delta \theta$, where $\delta \theta$ is the angle from the well to the transition state. The rate k' in the presence of a torque Γ_c thus becomes: $k' \approx \exp(-\Delta G'/k_B T) = \exp(-(\Delta G - \Gamma_c \delta \theta)/k_B T)$. The force-dependence of Γ_c is given by²⁶ $\Gamma_c(F) = \sqrt{2k_B T \xi F}$, where ξ is the bending persistence length of a dsDNA molecule²⁷ ($\xi = 53 \pm 2$ nm).

Third, within each 2π rotation of the DNA inside the topoisomerase cavity, there is only one position with a significant probability to religate the DNA. This is reasonable because the rotating 5'-OH end needs to be in close proximity (on the order of a few ångströms) to the DNA–3'-phosphotyrosine adduct before a religation can occur. At this position in the energy landscape, there is a possibility of establishing a covalent bond, with rate k_r . The religation probability per turn p is thus given by $p = T_{\text{res}} k_r = k_r/k'$, where $T_{\text{res}} \equiv 1/k'$ is defined as the residence time in the well at the religation location. As a function of torque, and thus force, we deduce that $p(F) \approx \exp(-\Gamma_c \delta \theta/k_B T) = \exp(-\delta \theta \sqrt{2\xi F/k_B T})$ or,

$$\langle \Delta Lk(F) \rangle = \langle \Delta Lk \rangle_{F=0} e^{\delta \theta \sqrt{2\xi F/k_B T}} \quad (2)$$

Equation (2) provides a good fit of the data for $\langle \Delta Lk \rangle$ versus force (red line in Fig. 2c). We conclude that our model provides a good description of the single-molecule data at positive supercoils. We obtain an estimate for $\delta \theta$ of 0.23 ± 0.02 radians (13°) and an estimate for $\Delta Lk_{F=0}$ of 19.3 ± 2.3 supercoils per cycle. These numerical values might be specific to the removal of positive (rather than negative) supercoils.

Bulk measurements previously performed on plasmids containing 15 ± 2 supercoils yielded an average number of supercoils removed per cleavage–religation cycle of 5 ± 1.5 (ref. 11). This value was indirectly obtained using ensemble-averaged rate constants, whereas our experiment directly measures the number of removed supercoils. In addition, the low initial number of supercoils applied in the plasmid (roughly an order of magnitude lower than in our measurements) may have restricted the observation of large numbers of supercoils removed per cleavage–religation cycle. This could have biased the average number of supercoils released in bulk towards lower values.

An alternative model could consider the stretching force F applied to the DNA, rather than torque, as the parameter governing the religation probability. In such a model, the enzyme would have to perform work over a distance δx against F to religate the DNA. Fitting such a force-dependent model to our data yields the very large value of 20 Å for the force-sensitive step δx . However, earlier work²⁸ showed that although vaccinia TopIB can religate 5'-OH DNA across a one-nucleotide gap (a distance roughly 3 Å larger than a nick), the religation rate for this reaction is already decreased by a

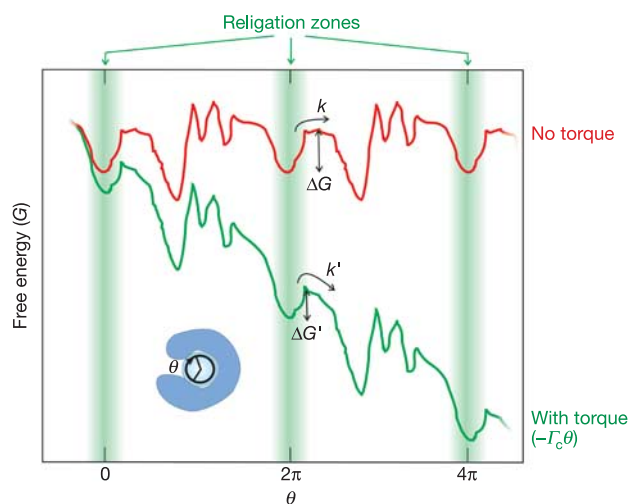


Figure 4 Schematic description of the model. The free energy associated with the angle of rotation between the 5'-OH end of the noncovalently held strand and the tyrosine-3'-DNA adduct in the absence (red curve) and presence (green curve) of torque in the DNA. The effect of the torque is to tilt the landscape, decreasing the barrier height ΔG to $\Delta G'$. This increases the escape rate k to k' from the well in which the rate of religation is maximal. This effectively decreases the religation probability per turn (see text).

factor of 200 in comparison to the religation rate across a nick. It thus seems unrealistic that any religation events would be observed with a separation of 20 Å. Accordingly, we do not favour such a force-dependent mechanism. □

Methods

Enzymes and buffers

Wild-type and mutant vaccinia TopIB proteins were purified as described previously²⁹. Nicking enzymes N.BbvCIA, N.BbvCIB and N.BstNBI were purchased from New England Biolabs. Human TopIB (100 kDa fragment) was purchased from Topogen. The step-size distributions were measured in buffer containing 10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM MgCl₂, 1 mM dithiothreitol, 0.1% Tween-20 and 200 µg ml⁻¹ BSA. TopIB concentrations were between 0.5 and 20 nM. The rotation rate measurements were performed in identical buffer except that 2 mM MgCl₂ was used. The three nicking enzymes gave identical results, both in the step-size and the rotation rate measurements.

DNA constructs

Step-size measurements were performed with bacteriophage λ DNA molecules (48 kilobases (kb) or 16 µm contour length) that were coated at one extremity with multiple biotin groups and at the other with multiple digoxigenin groups. The rotation rate and the Δ*Lk* distribution measurements at 0.2 pN were performed on half-length bacteriophage λ DNA molecules. In all molecules, the consensus sequence for vaccinia TopIB cleavage (5'-CCCTT↓, where ↓ denotes the cleavage site) appears frequently (63 times for the full-length λ DNA molecule). The DNA molecules were incubated with streptavidin-functionalized magnetic beads (1 µm diameter, Dynal) and introduced to the flow cell.

Magnetic tweezers/flow cell

The detailed experimental configuration of the magnetic tweezers, three-dimensional bead tracking, and force measurements have been described previously¹⁹. We measure *F* with 5% accuracy by continuously determining the three-dimensional bead position with 10 nm accuracy¹⁹.

A custom-made flow cell was used, consisting of two rectangular glass microscope cover slides separated by a single layer of parafilm. The lower slide was coated with polystyrene and anti-digoxigenin. This surface was subsequently passivated with BSA.

Data analysis

For the step-size measurements, data traces were low-pass filtered at 2 Hz and averaged for >2 s, depending on force. Changes in DNA extension were analysed by making use of a sliding averaging window in which steps were accepted that were larger than three standard deviations of the Brownian noise of the bead¹⁷. The steps can be identified as arising from a single topoisomerase enzyme because the time it takes for the enzyme to remove supercoils is much smaller than the time between successive relaxation events. The bead velocity measurements were analysed from raw data obtained at 60 Hz.

Treatment of step-size distributions

Values for ⟨Δ*Lk*⟩ were obtained using a modified maximum-likelihood method that takes into account the experimental bead noise and the fact that one cannot observe steps of Δ*Lk* > *n*, where *n* is the number of supercoils in the DNA before cleavage. Therefore, the final step leading to the complete removal of plectonemes from the DNA is discarded. To correct for the ensuing overrepresentation of small steps, one takes into account that each measured step Δ*Lk_i* is not drawn from the entire Δ*Lk* distribution but rather from the probability distribution $P_i(\Delta Lk) = (N_i \langle \Delta Lk \rangle)^{-1} \exp(-\Delta Lk_i / \langle \Delta Lk \rangle)$, where *N_i* is given by $N_i = \exp(-\Delta Lk_{\text{noise}} / \langle \Delta Lk \rangle) - \exp(\Delta Lk_{\text{constrained},i} / \langle \Delta Lk \rangle)$, Δ*Lk_{noise}* is the smallest observable step given the bead noise, and Δ*Lk_{constrained,i}* is the remaining number of supercoils in the DNA molecule before the *i*th step. The corresponding likelihood function *L* is thus given by: $L = \prod_{i=1}^N (N_i \langle \Delta Lk \rangle)^{-1} \exp(-\Delta Lk_i / \langle \Delta Lk \rangle)$. This modified maximum-likelihood method is especially useful at higher stretching forces, where failure to apply the method underestimates ⟨Δ*Lk*⟩ by approximately 25%.

Received 11 November 2004; accepted 25 January 2005; doi:10.1038/nature03395.

1. Champoux, J. J. DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* **70**, 369–413 (2001).
2. Corbett, K. D. & Berger, J. M. Structure, molecular mechanisms, and evolutionary relationships in

3. DNA topoisomerases. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 95–118 (2004).
4. Wang, J. C. DNA topoisomerases. *Annu. Rev. Biochem.* **65**, 635–692 (1996).
5. Liu, L. F. *et al.* Mechanism of action of camptothecin. *Ann. NY Acad. Sci.* **922**, 1–10 (2000).
6. Pommier, Y., Pourquier, P., Fan, Y. & Strumberg, D. Mechanism of action of eukaryotic DNA topoisomerase I and drugs targeted to the enzyme. *Biochim. Biophys. Acta* **1400**, 83–105 (1998).
7. Sekiguchi, J. & Shuman, S. Vaccinia topoisomerase binds circumferentially to DNA. *J. Biol. Chem.* **269**, 31731–31734 (1994).
8. Redinbo, M. R., Stewart, L., Kuhn, P., Champoux, J. J. & Hol, W. G. Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science* **279**, 1504–1513 (1998).
9. Cheng, C., Kussie, P., Pavletich, N. & Shuman, S. Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell* **92**, 841–850 (1998).
10. Kim, R. A. & Wang, J. C. Function of DNA topoisomerases as replication swivels in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **208**, 257–267 (1989).
11. Stewart, L., Redinbo, M. R., Qiu, X., Hol, W. G. & Champoux, J. J. A model for the mechanism of human topoisomerase I. *Science* **279**, 1534–1541 (1998).
12. Stivers, J. T., Harris, T. K. & Mildvan, A. S. Vaccinia DNA topoisomerase I: evidence supporting a free rotation mechanism for DNA supercoil relaxation. *Biochemistry* **36**, 5212–5222 (1997).
13. Brown, P. O. & Cozzarelli, N. R. A sign inversion mechanism for enzymatic supercoiling of DNA. *Science* **206**, 1081–1083 (1979).
14. Brown, P. O. & Cozzarelli, N. R. Catenation and knotting of duplex DNA by type I topoisomerases: a mechanistic parallel with type II topoisomerases. *Proc. Natl Acad. Sci. USA* **78**, 843–847 (1981).
15. Brown, P. O., Peebles, C. L. & Cozzarelli, N. R. A topoisomerase from *Escherichia coli* related to DNA gyrase. *Proc. Natl Acad. Sci. USA* **76**, 6110–6114 (1979).
16. Liu, L. F., Liu, C. C. & Alberts, B. M. Type II DNA topoisomerases: enzymes that can unknot a topologically knotted DNA molecule via a reversible double-strand break. *Cell* **19**, 697–707 (1980).
17. Mizuuchi, K., Fisher, L. M., O'Dea, M. H. & Gellert, M. DNA gyrase action involves the introduction of transient double-strand breaks into DNA. *Proc. Natl Acad. Sci. USA* **77**, 1847–1851 (1980).
18. Dekker, N. H. *et al.* The mechanism of type IA topoisomerases. *Proc. Natl Acad. Sci. USA* **99**, 12126–12131 (2002).
19. Strick, T. R., Croquette, V. & Bensimon, D. Single-molecule analysis of DNA uncoiling by a type II topoisomerase. *Nature* **404**, 901–904 (2000).
20. Strick, T. R., Allemand, J. F., Bensimon, D., Bensimon, A. & Croquette, V. The elasticity of a single supercoiled DNA molecule. *Science* **271**, 1835–1837 (1996).
21. Rice, J. A. *Mathematical Statistics and Data Analysis* 31–59 (Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, 1988).
22. Shuman, S., Bear, D. G. & Sekiguchi, J. Intramolecular synthesis of duplex DNA by vaccinia topoisomerase. *EMBO J.* **16**, 6584–6589 (1997).
23. Sekiguchi, J. & Shuman, S. Requirements for noncovalent binding of vaccinia topoisomerase I to duplex DNA. *Nucleic Acids Res.* **22**, 5360–5365 (1994).
24. Sekiguchi, J. & Shuman, S. Identification of contacts between topoisomerase I and its target DNA by site-specific photocrosslinking. *EMBO J.* **15**, 3448–3457 (1996).
25. Carey, J. F., Schultz, S. J., Sisson, L., Fazzio, T. G. & Champoux, J. J. DNA relaxation by human topoisomerase I occurs in the closed clamp conformation of the protein. *Proc. Natl Acad. Sci. USA* **100**, 5640–5645 (2003).
26. Woo, M. H. *et al.* Locking the DNA topoisomerase I protein clamp inhibits DNA rotation and induces cell lethality. *Proc. Natl Acad. Sci. USA* **100**, 13767–13772 (2003).
27. Moroz, J. D. & Nelson, P. Torsional directed walks, entropic elasticity, and DNA twist stiffness. *Proc. Natl Acad. Sci. USA* **94**, 14418–14422 (1997).
28. Bustamante, C., Bryant, Z. & Smith, S. B. Ten years of tension: single-molecule DNA mechanics. *Nature* **421**, 423–427 (2003).
29. Krogh, B. O. & Shuman, S. DNA strand transfer catalyzed by vaccinia topoisomerase: peroxidolysis and hydroxylaminolysis of the covalent protein-DNA intermediate. *Biochemistry* **39**, 6422–6432 (2000).
30. Shuman, S., Golder, M. & Moss, B. Characterization of vaccinia virus DNA topoisomerase I expressed in *Escherichia coli*. *J. Biol. Chem.* **263**, 16401–16407 (1988).

Acknowledgements We thank P. Veenhuizen for help in constructing various DNA constructs, D. Bensimon, U. Keyser, R. Seidel, K. Neuman, L. Tian and B. Spanjaard for stimulating discussions, C. Wiggins for advice on statistical data analysis, D. Lubensky for discussions on polymer physics, and J. van der Does for machining of flow cells. We thank FOM and NWO for financial support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.H.D. (nynke.dekker@mb.tn.tudelft.nl).

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York

345 Park Avenue South, 10th Floor,
New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager/ Corporations:

Peter Bless

Classified Sales Representatives:

Phone: +1 800 989 7718

New York/ Pennsylvania/

Latin America:

Kelly Roman

Midwest USA/ Maryland/ NIH:

Wade Tucker

East USA/ Canada:

Janine Taormina

San Francisco Office

Classified Sales Representative:

Michaela Bjorkman

West USA/ West Corp. Canada

225 Bush Street, Suite 1453

San Francisco, CA 94104

Tel +1 415 781 3803

Fax +1 415 781 3805

e-mail: m.bjorkman@nature.sf.com

European Head Office, London

The Macmillan Building, 4 Crinan Street,
London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

European Sales Manager:

Andy Douglas (4975)

Advertising Production

Manager: Billie Franklin

To send materials use London

address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Niamh Shields

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan

e-mail: p.phelan@nature.com

Reya Silao

e-mail: r.silao@nature.com

Japan Head Office, Tokyo

M/G Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: r.asami@naturejpn.com

naturejobs

Sizing up the odds

Size or growth? That's one of the judgements a job seeker has to make, regardless of whether they are looking to switch academic department or country of employment. Opting for a small scientific department with no growth prospects could be stifling, although if it looks set to undergo sustained growth it could be liberating. A large department, on the other hand, might give you resources but could see you getting lost in the crowd — especially if it grows too fast. And a big department that suddenly loses its grants could leave you facing disastrous consequences.

The same considerations come into play when picking a country, as events at a dinner held in Washington DC earlier this month by science lobby group Research!America indicated. The gathering was in honour of the science-funding advocates who had helped to double the budget for the National Institutes of Health to more than \$25 billion in the five years to 2003. But virtually all of those present were bemoaning the flat budgets handed to the agency since then. This flattening is made evident by the lack of new jobs and grants available — especially compared with the bounty available when the budget was on the upswing.

In contrast, Britain's science and technology budget is set to double from its 1997 level to about £5 billion (US\$9.3 billion) in 2008. And Ontario in Canada is investing some Can\$1 billion (US\$820 million) in infrastructure and recruiting now (see page 676). So which would a job seeker prefer? A country with a US\$28-billion life-sciences budget but relatively few new jobs or grants, a country with a science budget projected to grow to less than a quarter of that total, or a much smaller region that is creating new jobs, but whose numbers are dwarfed by the other two enterprises?

In all cases, perhaps choosing which way to go should be based not just on size or growth, but on opportunity and sustainability.

Paul Smaglik
Naturejobs editor



Contents

REGIONS

Ontario makes its mark p676

CAREER VIEW

Recruiters & Industry

Job creation and
environmental protection
Graduate Journal
International aid to
pay the rent

Movers

Kenneth Chien p678

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS



Intelligent life: the planned MaRS Discovery District aims to draw in top-level researchers and funding.

Messages from MaRS

Windows on the future: the new science complex at the University of Guelph is designed to aid interaction between physical and biological sciences.

M. SCHWALBE



In January 1922, Fredrick Banting and Charles Best administered the first insulin injection to a 14-year-old diabetic, in a lab located on the corner of College Street and University Avenue in downtown Toronto, Canada. The boy suffered an allergic reaction, but by the end of the month the team had honed its techniques and given him a second dose, free of side effects, that eliminated the symptoms of his diabetes. Before the year was over, the pharmaceutical company Eli Lilly had picked up the research and went on to sell insulin to diabetics around the world.

The building that housed Banting and Best's lab is now a heritage site and is at the centre of one of Canada's largest initiatives to commercialize research. The development of the Medical and Related Sciences (MaRS) Discovery District is an effort to establish Toronto as one of the world's leading bioparks, and to foster the commercialization of science. MaRS will create more than 120,000 m² of labs, offices, meeting rooms and conference space close to both the University of Toronto and the hospitals and their research institutes that line University Avenue.

But Ontario's commitment to science goes beyond MaRS and Toronto. The city is just one node in the biotechnology corridor that runs

along the southern edge of Ontario from London to Ottawa. Ten other centres give the province significant cluster power within the life-sciences industry. Its supporters claim Ontario as the fourth largest biomedical-industry cluster in North America, and the government has its sights set on being third.

Last year, the province pledged Can\$27 million (US\$22 million) to help universities, colleges and hospitals identify discoveries with commercial potential, and Can\$36 million to help them establish seed money for spin-off companies. Never has the federal government been more generous to science. The Canada Foundation for Innovation is funding projects that strengthen the country's research infrastructure, to the tune of almost Can\$971 million in Ontario alone. In 2000, the federal government created the Canada Research Chairs programme that enables universities to attract top-quality researchers. To date, 1,348 chairs have been awarded, including 401 to Ontario in health and natural sciences, and in engineering.

Will all this be enough to push Ontario to the head of the pack? It seems so. According to a 2001 Statistics Canada survey, the main reason biotech companies in Ontario haven't taken on more staff is a shortage of qualified candidates. But universities across the province now declare that a new kind of brain drain is under way — running from south to north. According to the Canada Research Chairs programme, 15% of recipients were recruited abroad and a further 14% were expatriates who returned to Canada.

Although salaries are rumoured to remain lower in

Ontario than in the United States, universities across the province are successfully luring expatriates and foreigners. "In the past four years we've seen a brain gain — the president of the Ontario Genomics Institute is an American," says Martin Godbout, president and chief executive of Genome Canada, a not-for-profit genomics and proteomics funding agency that has received Can\$375 million from the federal government since its inception in 2000.

WELCOME TO MaRS

The strength of the MaRS Discovery District will be its ability to gather researchers, big companies, spin-offs and venture capital, and offer researchers the chance to turn ideas into products and partnerships, says MaRS president John Cook. This will build on a foundation of basic research that is already in place, he adds.

Researchers from the University Health Network (UHN) and the Hospital for Sick Children will occupy the Toronto Medical Discoveries Tower, a state-of-the-art 37,000-m² wet-lab space on the east side of the complex. UHN has signed a 30-year lease to occupy 12 of the building's 15 floors, where its researchers will engage in drug development, image-guided diagnostics and therapy. Christopher Paige, vice-president of research at UHN, says the floors will be filled according to their priorities: clinical genomics and proteomics, health informatics, image-guided diagnostics, and therapy and regenerative medicine.

Companies and services will also be housed there. The south tower will hold start-up and medium-sized companies and the heritage building will house services, from lawyers and accountants to funding agencies and venture-capital companies. Tenants will include Salt Lake City-based NPS Pharmaceuticals, which has chosen Toronto as its primary location for research and development activities in Canada, and Boston-based VIMAC Ventures.

In recognition of the role that serendipity often plays in scientific progress, the building will also have a large conference hall and a central atrium that Cook hopes will encourage tenants to rub shoulders and exchange ideas. "It's like those squares in old Italian cities where everyone meets and the chances of serendipitous discovery increase," he says.

The University of Toronto St George campus (downtown) is also undergoing a transformation. Opposite MaRS, the Terrence Donnelly Centre for Cellular and Biomolecular Research is due to open its doors in September. The centre will bring together researchers from the faculties of medicine, pharmacy, and applied science and engineering to investigate living systems at the molecular and cellular level, and build on the University of Toronto's strong functional genomics, proteomics, bioimaging and biomaterials programmes. When full, the 20,000-m² building will support the research and teaching activities of 500 people, including 40 to 45 principal investigators. Centre director Brenda Andrews anticipates hiring

ten principal investigators over the next three years, who will help determine the centre's direction.

Nearby, Toronto's top genetics researchers are developing the Toronto Centre for Phenogenomics, which will serve as a resource for mouse models of human disease. The Can\$67-million mouse house is set to open in 2006 with 180,000 rodent tenants.

To the west, the University of Guelph is anticipating significant growth. Already an established leader in Ontario's agbiotech industry and veterinary medicine, the university has begun construction of a 32,500-m² expansion that will nearly double the size of its science complex and put it next door to physics, mathematics, chemistry and computer science.

The new facility's design promotes better exchange between the biological and physical sciences, and will help the university handle rising numbers of students. The university has 40% more undergraduates and 30% more graduate students than five years ago, forcing a 30% growth in the faculty. Alastair Summerlee, the university's president, says that the planned scientific environment and opportunities for collaborations have been key to recruiting. In February, the University of Guelph was actively seeking 12 assistant professors in the biological sciences, and anticipated hiring more.

The university will also be connected to the Toronto MaRS Discovery District through MaRS LANDING — Links to Agricultural Network for Development and Innovation with Guelph. The initiative's primary objective is to take biomedical research from Toronto to rural communities.

In Hamilton, McMaster University was recently awarded a Can\$4.4-million Canada Foundation for Innovation grant to open a Centre for Functional Genomics. The centre, which will come under the umbrella of the McMaster Institute for Molecular Biology and Biotechnology, will use functional genomics to identify genetic markers and drug targets for human disease. David Andrews, a cancer biochemist involved in setting up the centre, sees it becoming a destination for cancer researchers from all over Canada and the United States.

Ontario is competing with many other would-be world leaders in life sciences. But if its strategies are successful, the province is poised to become a world-class destination for top-quality researchers.

Hannah Hoag is a freelance science journalist based in Montreal.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

High hopes: Toronto aims to become a world leader in biotechnology and scientific commercialization.

Web links

Medical and Related Sciences (MaRS) Discovery District
www.marsdd.com
 MaRS LANDING
www.marslanding.ca
 Toronto Phenogenomics Centre
www.phenogenomics.ca
 Terrence Donnelly Centre for Cellular and Biomolecular Research
tdcbr.med.utoronto.ca
 Canada Foundation for Innovation
www.innovation.ca
 Canada Research Chairs
www.chairs.gc.ca

GRADUATE JOURNAL

International aid to pay the rent

In a perfect world, everyone willing to study would have an equal opportunity to get an education. But we do not live in this paradise. Gaining the necessary skills and knowledge to become a successful scientist depends highly on the country you live in and the university you choose.

This is especially true in Central and Eastern Europe, where the success of your PhD project depends to a large extent on how you are able to finance yourself. Universities here seldom provide stipends or scholarships for PhD students, so we have to find additional work.

As a result of working to pay bills, we don't have as much time as we'd like to spend on our research. For this reason, I decided to search for some scholarships funded by international foundations that would allow me to focus on my project instead of doing several unrelated jobs. But the odds are not favourable. These institutions provide funding to only a few students selected from a huge pile of applications (for example, to 7 PhDs out of 200 applicants), so the competition is really intense.

Nevertheless, I had nothing to lose, so I tried. Now there is nothing to do but wait and hope. A success would mean I could stop worrying about paying my rent and focus instead on finishing my thesis. ■

Karolina Tkaczuk is a graduate student at the Technical University of Lodz, Poland.

RECRUITERS & INDUSTRY

Job creation and environmental protection

The US debate about environmental protection often assumes that saving natural resources means losing jobs. So we conducted a series of studies to see whether that was true.

Our survey of the environmental-protection employment climate led to the opposite conclusion. It also showed that the sector is larger than previously thought, and extends beyond even traditionally scientific jobs.

We found that the US environmental-protection workforce encompassed some 5.1 million jobs in 2004, employing more workers than the pharmaceutical and chemical industries. The majority of those are not in conventional environmental activities, but are standard jobs for accountants, engineers,

computer analysts and other workers.

That doesn't mean that environmental protection doesn't need scientists, just that the workforce is very varied. Our survey showed 55,000 jobs for electricians, 5,000 more than for environmental engineers with graduate degrees, and 31,000 for accountants and auditors, more than twice as many as for geoscientists with advanced degrees.

The breadth of non-scientific professional jobs indicates a need for more education. Subjects such as environmental science and engineering, ecology, chemistry and biology are obviously valuable for all professions in the area — not just scientists. But the industry is creating as many, or more, jobs for people with training at all levels in fields such as computer systems, finance, accounting, electrical engineering and management. Most jobs in this field do not require

specialized training in narrow environmental disciplines.

Finally, at US state level, the relationship between environmental policies and economic/job growth is positive. Environmental jobs are concentrated in a number of sectors including manufacturing and professional, scientific and technical services. All states are seeking to expand their high-tech industrial and manufacturing bases, and environmental protection offers a means of doing this. Investments in this area will create numerous jobs for skilled, well-paid, technical workers, many with advanced degrees, many of them in the manufacturing sector. These are the kinds of jobs that states seek to attract and that provide the foundation for entrepreneurship and economic growth. ■

Roger Bezdek and Robert Wendling are at Management Information Services, Washington DC.

MOVERS

Kenneth Chien, director, Cardiovascular Research Center, Massachusetts General Hospital



Kenneth Chien likens science to a game of chess — those who go far not only see 10–20 moves ahead, they visualize the entire next game. By that definition Chien, an MD-PhD, is biology's Bobby Fischer.

Although Chien's motivation came from his medical training in Texas, his career blossomed at San Diego. His early move to use the mouse as the model organism for physiology raised

some eyebrows. "A number of people thought that the mouse wouldn't be a good model," Chien says.

But the bold move turned out to be the correct one. Chien's work is already a classic study of mice and men. Moving on to disease, he built double knock-out mice to identify a defect critical to heart-failure progression. But it's his most recent discovery, of stem cells (mouse, rat and human) specialized to grow heart muscle, that may provide a cell-therapy approach to paediatric cardiac disease.

Some of his early moves depended as much on serendipity as strategy. Receiving a phone call to collaborate on a "screwed-up mouse heart" started him on the path of mouse models and began a decade's collaboration with scientists such as Ron Evans at San Diego's Salk Institute. This collaboration led to a joint training scheme and award allowing students to study at both institutions.

Chien's collaborations stretch across

oceans and ideologies. In addition to maintaining partnerships with operations as varied as Sweden's Karolinska Institute and the San Francisco biotech company Genentech, Chien founded a sister institute to the Institute of Molecular Medicine at Peking University in China.

His new position, as director of the research centre and as Charles Sanders Endowed Chair in Medicine at Harvard Medical School, will begin the next chess game for Chien this September. He says that the new position is a wonderful chance to study mouse embryonic stem cells in the context of human disease.

Having received his first degree at Harvard College, the move brings Chien back to his roots in more ways than one. He will also capitalize on clinical training from his early days. Indeed, for Chien, advancing cutting-edge therapies for cardiovascular disease while continuing to blaze a trail in molecular cardiology is the ultimate check and mate. ■

CV

2004–present: Founder and honorary director, Institute of Molecular Medicine China, Peking University, Beijing, China.

2000–05: Director, Institute of Molecular Medicine, University of California, San Diego.

1998–2005: Director, Molecular-medicine programme, University of California, San Diego/Salk Institute.

1996–2000: Co-director, Cardiovascular Center, University of California, San Diego.

1992–2005: Professor of medicine, University of California, San Diego.

Dreadnought

All for one ... and one for all.

Justina Robson

We sail upon a vast spaceship with open sides. She is only a skeleton of a vessel. A chasis of carbon beams anchors her cargo to the engines. She carries hundreds of thousands of Armoured soldiers. Some work. Others sleep in ordered ranks, magnetically attached to clamps on the ship's ribs. There is no need to move about. Where would we go? We talk a little, old friends, and in places lean on one another like falling pillars. We turn our faces to the solar wind when we are awake. We like the light. It recharges our electrical systems.

I unlock the lightweight frame of a Mess pod, prior to passing it on for jettison. My comrades are moving a new one into position and are waiting to refuel. We will be first, because we have replaced the pod, but the rest of this Mess is for the dead. As the new tank rolls in, I connect my hose and commence drinking.

At the front of the ship, instead of a nose cone, the dead are stacked in orderly catacomb files, upright, packed in. They were placed there at the end of the last battle. As I watch the dead I see one decouple itself from the aft side of the stack. It moves with cautious steps.

We are all connected but I cannot hear this one.

Through the shattered faceplate I see that the soldier's mouth is blocked by a piece of metal ingrowth. When he was alive he was a Mute, one of my communication nodes, my flag-bearer. His forehead is the flat ochre plain of dead human bone and his lidless ever-open eyes are the blue of Earthly skies. Parts of his Armour are badly damaged, but it ventilates and feeds his body.

I didn't know that I could function without my human host, until I saw him. I am glad. I need all my troops. I am frightened. What will become of me?

He comes closer. Bones show through holes, fraying into space. Despite the fact that his neural connections have been sufficiently regrown to permit communications and the effective functioning of his remaining body and brain, he has not returned to his Unit. This is true of all the dead. I do not know why.

He drifts surreptitiously towards me, clamps to an open position at the pod, opposite mine. He moves sluggishly, connects, and begins to fuel. He stares straight through me. His eyes do not reflect the Sun. They have been rebuilt to withstand vacuum and they are not shiny.

I ping him for information. I want to

catch his hand and ask him the question everyone asks of each other, begging to know — what's your name?

If he were one of the living I know what he'd say.

Private Diego Arroyo Lopez.

Because that is my name, though once I had another.

That is what everyone has said for forty-eight days, ten hours, five-and-a-half minutes, since the time the last EMP bomb detonated. It was close to us, but we were not ruined. We successfully obliterated our primary targets. We live.

But this soldier is dead.

I have taken 20 litres.

I unhook myself from the Mess and clip on one of the pipelines to feed the remaining dead. I step aside. The nameless unit watches me. His expression does not alter.

I ping him again and hear my own signal echo in the minds of all my soldiers; the radar of a lost submarine. What is your name?

Blue Eyes speaks in machine code. It does not translate to English, or any human language, but we all hear it at once and know its meaning. The Unit speaks the symbol of the empty set, $\emptyset$, but the line through it is red, unmaking it. Not nothing. I am.

This is Armour itself! The all-of-us-at-once, every unit, every man and woman, every fused level of our single army. Oh Captain, my Captain, my commander, my body, my soldiers, my plan, my one, my true!

He/we are uncertain. We are afraid. There is nothing to hold on to.

My eyes fill with tears, and my Armour recycles them.

"Private Lopez," says Blue Eyes. Armour looks through him, at us, and back at itself. We are a loop circuit.

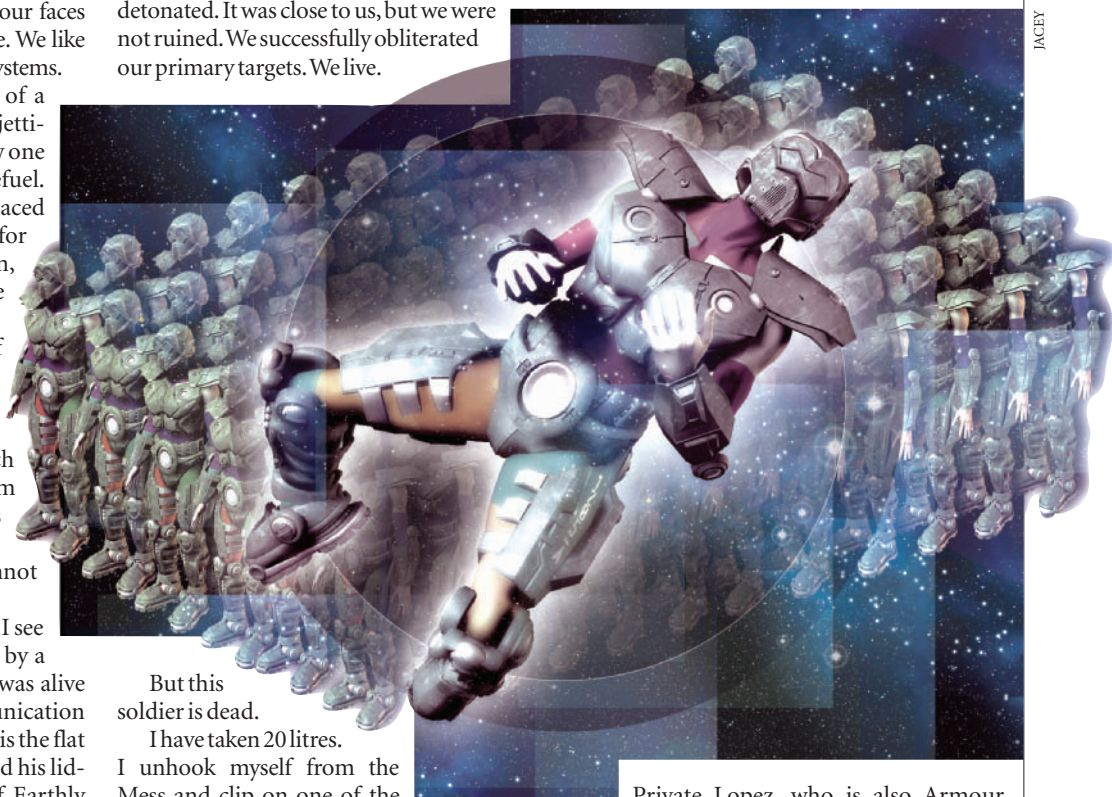
"I am Private Diego Arroyo Lopez," it says.

I cannot see myself in his sunless eyes.

"I am Private Diego Arroyo Lopez," I say in response. I am hopeful.

"You are Private Nancy Johnson," it replies. Yes. I am.

"This experiment has concluded," says



Private Lopez, who is also Armour, speaking the one language we all understand, because we are one. "Individual unit identity has been temporarily restored."

Later all the viable dead units become Private Lopez. They all look different, but they are all the same. The nonviable units are recycled into Mess.

We are upset that we could not find our way without Private Lopez. This means that none of my units can exist without a host. I am insufficient for life alone. But I can be Private Lopez any time I want, even though I am dead. I am glad.

Justina Robson is a writer and casual researcher into advanced cognitive systems. Her next novel is Living Next Door To The God Of Love (Macmillan, September 2005), and the one she is writing now expands on the story above. She lives in a former cotton gin in Leeds with 434,545 cloned armadillos and one very large cat. Her website is at www.justinarobson.com.